Analyzing cellular immunogenicity in vaccine clinical trials: a new statistical method including non-specific responses for accurate estimation of vaccine effect

Édouard Lhomme, Boris Hejblum, Christine Lacabaratz, Aurélie Wiedemann, Jean-Daniel Lelièvre, Yves Lévy, Rodolphe Thiébaut, Laura Richert

Analyzing cellular immunogenicity in vaccine clinical trials: a new statistical method including non-specific responses for accurate estimation of vaccine efficacy

Edouard Lhomme[1,2,3,4], Boris P. Hejblum[1,2], Christine Lacabaratz[2,5], Aurélie Wiedemann[2,5], Jean-Daniel Lelièvre[2,5], Yves Levy[2,5], Rodolphe Thiébaut[1,2,3,4], Laura Richert[1,2,3,4]

[1] Univ. Bordeaux, Department of Public Health, Inserm Bordeaux Population Health Research Centre, Inria SISTM, F-33000 Bordeaux, France
[2] Vaccine Research Institute (VRI), Créteil, F-94000 France
[3] Pôle de Santé Publique, CHU de Bordeaux, Bordeaux, F-33000 France
[4] Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219 ; CHU Bordeaux ; CIC 1401, EUCLID/F-CRIN Clinical Trials Platform, F-33000, Bordeaux, France
[5] Inserm U955, Henri Mondor Hospital, University of Paris East, F-94000 Créteil, France

Corresponding author: Edouard Lhomme & Rodolphe Thiébaut
Permanent address:
ISPED – Université de Bordeaux
146 rue Léo Saignat
33076 Bordeaux Cedex
France

Abstract

Evaluation of immunogenicity is a key step in the clinical development of novel vaccines. T-cell responses to vaccine candidates are typically assessed by intracellular cytokine staining (ICS) using multiparametric flow cytometry. A conventional statistical approach to analyze ICS data is to compare, between vaccine regimens or between baseline and post-vaccination of the same regimen depending on the trial design, the percentages of cells producing a cytokine of interest after *ex vivo* stimulation of peripheral blood mononuclear cells (PBMC) with vaccine antigens, after subtracting the non-specific response (of unstimulated cells) of each sample. Subtraction of the non-specific response is aimed at capturing the specific response to the antigen, but raises methodological issues related to measurement error and statistical power. We describe here a new statistical approach to analyze ICS data from vaccine trials.

We propose a bivariate linear random-effect regression model for estimating the non-specific and antigen-specific ICS responses. We benchmarked the performance of the model in terms of both bias and control of type-I and -II errors in comparison with conventional approaches, and applied it to simulated data as well as real pre- and post-vaccination data from two recent HIV vaccine trials (ANRS VRI01 in healthy volunteers and therapeutic VRI02 ANRS 149 LIGHT in HIV-infected participants).

The model was as good as the conventional approaches (with or without subtraction of the non-specific response) in all simulation scenarios in terms of statistical performance, whereas the conventional approaches did not provide robust results across all scenarios. The proposed model estimated the T-cell responses to the antigens without any effect of the non-specific response on the specific response, irrespective of the correlation between the non-specific and specific responses.

This novel method of analyzing T-cell immunogenicity data based on bivariate modelling allows consideration of all T-cell data and is more flexible than conventional

methods, and so yields more detailed results and enables accurate interpretation of vaccine efficacy.

Keywords: Flow cytometry; intracellular cytokine staining; vaccine; immunogenicity; clinical trials

Highlight

- Evaluation of vaccine immunogenicity is a key step in the clinical development of vaccines; the T-cell responses to vaccine candidates are typically assessed by intracellular cytokine staining using flow cytometry.
- Conventional approaches for analyzing T-cell responses can bias estimates and compromise statistical performance, particularly in terms of the type-I error rate and statistical power.
- We propose a new modeling approach that considers all measured data and is more flexible than conventional methods, which yields more detailed results and enables accurate interpretation of vaccine efficacy.

1. Introduction

In clinical development of vaccines, assessing the ability of the candidate vaccine to generate immune responses is an important objective of phase I and II clinical trials. Only candidate vaccines with sufficient immunogenicity are subjected to phase III clinical trials. Immunogenicity can be a secondary objective of phase III vaccine trials, to investigate potential correlates of protection.

In phase I and II vaccine trials, a variety of immunological markers, including the vaccine-induced T-cell responses, are typically assayed (1-4). T-cell responses to vaccine candidates are commonly assessed by intracellular cytokine staining (ICS), a flow cytometry assay involving cryopreserved peripheral blood mononuclear cells (PBMCs). ICS assay allows characterization of subsets of specific cytokine-producing T cells after *ex vivo* antigenic stimulation (5,7). For example, in the case of evaluation of an HIV vaccine candidate, the cellular responses can be measured by enumerating IFNγ–, IL2-, and TNFα-producing CD4+ and CD8+ T cells by flow cytometry, after stimulation of PBMCs with pools of the HIV peptides contained in the vaccine sequence, *e.g.*, Gag, Nef, Pol, and/or Env (8-11).

Different approaches exist for the statistical analysis of ICS data. Methods based on qualitative binary criteria can be used to report the proportion of vaccine recipients with a detectable or "positive" response (percentage of "responders"). The binary response variable can be defined using either empirical thresholds on a relative or absolute scale (*e*.g., positive response if higher than threefold the non-stimulated cells or at least 0.05% of T cells), relying on a statistical approach, or a combination of both. For instance a Fisher exact test on count data (number of flow cytometer events) is often used to compare the proportion of cytokine-producing stimulated and non-stimulated cells within a given sample, yielding a multiple testing-adjusted *p*-value per sample that is used as a statistical positivity criterion (5). However, in the absence of formally identified thresholds of positivity, the biological meaning of such binary criteria

remains unknown. In addition, a binary variable induces a loss of precision and information and therefore decreases the statistical power compared to using the full distribution of the marker (12). Qualitative criteria have the advantage of being easy to implement and interpret, without any positivity criteria to define. Thus, use of quantitative endpoints for the analysis of ICS responses is relevant. The gold standard (*i.e.*, conventional statistical approaches (1-4)) for the analysis of quantitative ICS data is: i) to subtract the response observed in non-stimulated cells from each stimulated condition of a given sample and ii) to perform a standard inter- or intra-arm comparison of the distribution of percentages of cells producing the cytokine(s) of interest. This can be achieved, for instance, using Student's *t*-test to compare two means of cytokine-producing cells at a given time point between trial arms (or for non-parametric data a rank test), or the corresponding tests for paired data for intra-group ("before-after immunization") comparisons. The background subtraction in step i) is aimed at capturing the antigen-specific response, but it can result in biased estimates and induce type-I errors (a type-I error occurs when a researcher rejects a null hypothesis when it is true), and reduce the statistical power by increasing the measurement error or biasing the observed distribution.

Moreover, although measurement errors and random biological variations are likely to contribute to the observed response in non-stimulated cells (often called "background"), it cannot be excluded that this non-specific response has some biological relevance (13,14), via bystander activation for instance, and should not be ignored in the data analysis. In addition, from a statistical point of view, the existence (or absence) of a correlation between non-specific and specific responses is significant (15). The conventional approaches do not take this into account, which may lead to erroneous results.

We propose a new statistical method for the analysis of cellular immune responses in vaccine trials using a bivariate regression model that guarantees accurate estimation of the vaccine effect.

2. Materials and Methods

2.1. Statistical model

We propose a bivariate linear model, which is an extension of the univariate regression models frequently used in biomedicine. While a univariate model allows estimation of the effects of one or several explanatory (independent) variables on a single response (dependent) variable, a bivariate model simultaneously includes two response markers as dependent variables (16). As for univariate linear regression models, a maximum-likelihood approach can be used to estimate the effects of the explanatory variables on the outcome in a bivariate model. Regression coefficients (so-called "betas") are thus estimated, along with their confidence intervals and corresponding *p*-values, while simultaneously modeling the vaccine effect on the non-specific and specific responses assayed by ICS.

The model considers the non-stimulated cell response and the raw stimulated cell response(s) as the dependent variables. These responses are modeled according to the vaccine effect as the main explanatory variable, and the stimulated cell response is additionally adjusted for the non-stimulated cell response. This adjustment allows consideration of the potential correlation between these responses, which is not possible with current conventional approaches. The model provides an estimation of the vaccine effect on the non-stimulated cell response and of the effect of the non-stimulated cell response on the stimulated cell response (Figure 1).

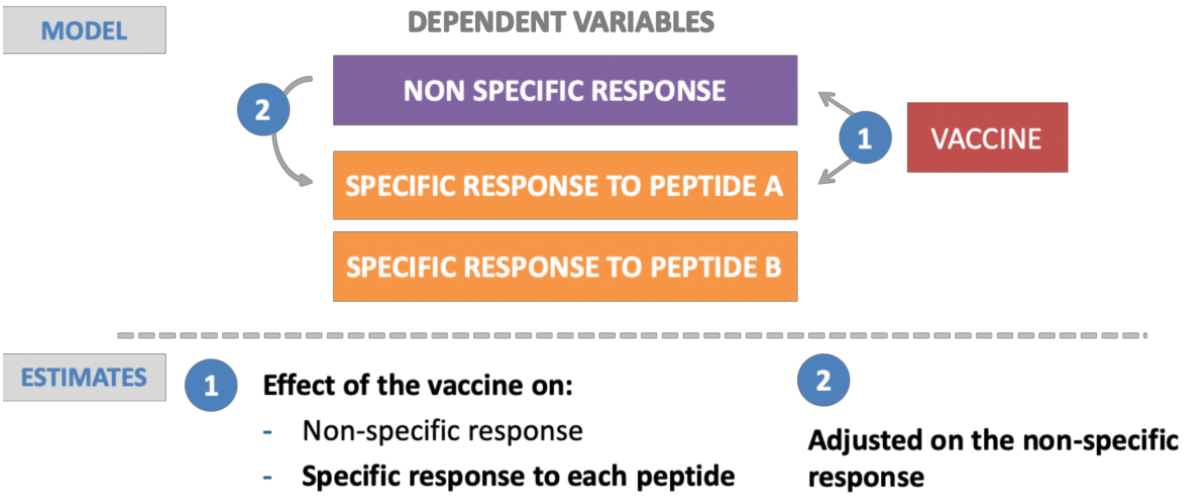


Figure 1. Bivariate linear model for estimating the non-specific and antigen-specific responses measured by ICS following stimulation *in vitro* with peptides A and B. The model is represented as a directed acyclic graph.

Within this framework, we explicitly propose two statistical models, depending on whether the vaccine trial is comparing multiple arms or not. The first model (1) is for a transversal between-trial arm comparison at a given time point, *i.e.*, in a comparative randomized trial comparing two vaccine regimens or an experimental vaccine *vs*. placebo. The second model (2) was developed for a within-arm comparison (comparison of post-vaccination *vs*. baseline for a single vaccine strategy), *i.e.*, non-comparative single- or multi-arm phase I/II vaccine trials of several strategies one by one (17,18).

The mathematical equations specifying each of the two models are described below.

Let $Y_i = \begin{bmatrix} Y_i^{NS} \\ Y_i^{S_k} \end{bmatrix}$, be the response vector for subject *i*, with Y_i^{NS} the non-stimulated cell response, and $Y_i^{S_k}$ the n^k vector of the stimulated cell response. We define the two bivariate linear models presented above as:

Model (1) for inter-arm comparison at one time point

$$\begin{cases} Y_i^{NS} = \beta_0^{NS} + \beta_1^{NS} V_i + \varepsilon_i^{NS} \\ Y_i^{S} = \beta_0^{S} + \beta_1^{S} V_i + \beta_2^{S} Y_i^{NS} + \varepsilon_i^{S} \end{cases}$$

where V_i is the variable indicating the vaccine arm, and β and ε, respectively, are the parameters and the errors of the model. This formulation implies the following probability distribution for the responses and errors:

$Y_i^{NS} \sim \mathrm{N}(\beta_0^{NS} + \beta_1^{NS} V_i, \sigma^{NS})$ and $\varepsilon_i^{NS} \sim N(0, \sigma^{NS})$

$Y_i^{S} \sim \mathrm{N}(\beta_0^{S} + \beta_1^{S} V_i + \beta_2^{S} Y_i^{NS}, \sigma^{S})$ and $\varepsilon_i^{S} \sim N(0, \sigma^{S})$

Model (2) for intra-arm comparison (post-vaccination vs. baseline)

$$\begin{cases} Ydiff_i^{NS} = \beta_0^{NS} + \varepsilon_i^{NS} \\ YdiffY_i^{S} = \beta_0^{S} + \beta_1^{S} Y_i^{NS} + \varepsilon_i^{S} \end{cases}$$

where $Ydiff_i^{NS} = Y_i^{NS}(T1) - Y_i^{NS}(T0)$ and $Ydiff_i^{S} = Y_i^{S}(T1) - Y_i^{S}(T0)$,$V_i$ β and ε are, respectively, the parameters and the errors of the model. This formulation implies the following probability distribution for the responses and errors:

$Ydiff_i^{NS} \sim \mathrm{N}(\beta_0^{NS} + \beta_1^{NS} V_i, \sigma^{NS})$ and $\varepsilon_i^{NS} \sim N(0, \sigma^{NS})$

$Ydiff_i^{S} \sim \mathrm{N}(\beta_0^{S} + \beta_1^{S} V_i + \beta_2^{S} Y_i^{NS}, \sigma^{S})$ and $\varepsilon_i^{S} \sim N(0, \sigma^{S})$

β_0^S We provide an implementation of the model in both SAS (using Proc Mixed) and R (using the nlme package). The SAS and R codes are provided in Appendix A. In addition, we built a user-friendly graphical interface that allows analysis of ICS data with the bivariate model and visualization of the results. The tool is implemented as an R-Shiny application and is available on the Internet (19).

2.2. Simulation study

Simulated datasets enable assessment of the performance of the proposed model in various scenarios where the truth is known (contrary to real-world data). Synthetic datasets were generated under various scenarios chosen to compare the behavior of the proposed model with the conventional approaches in different situations: a similar vaccine effect on stimulated cell response between arms to evaluate the risk of type-I error (*i.e.*, the risk of concluding that a

vaccine effect exists when it does not); the vaccine effect on the stimulated cell response to evaluate the statistical power (1-beta, or type II error); the presence and absence of correlations between the stimulated and non-stimulated responses; and with or without a vaccine effect on the non-stimulated response.

The various scenarios and parameters used to generate the data are described in Appendix B – Table 1. For each scenario, 1,000 simulations were run with three different sample sizes, respectively, 15, 30, and 60 participants per arm. Bias, type-I error control, and the statistical power of the bivariate model in each scenario were compared with those of two quantitative conventional approaches based on Student's *t*-test or paired *t*-test (for inter- and intra-arm comparisons, respectively) using (i) the raw stimulated cell response or (ii) the stimulated cell response after subtraction of the non-stimulated cell response. Relative bias was calculated as follows:

$$BIAS(\%) = \frac{DIFFobs - DIFFtheo}{DIFFtheo}$$

where *DIFFobs* is the observed mean of the difference between stimulated responses (between time points in model (1); between arms in f model (2)) and *DIFFtheo*, the corresponding theoretical difference (known in simulations). Type-I error was calculated as the percentage of simulations with significant vaccine efficacy ($p < 0.05$) among scenarios with no true vaccine efficacy. Statistical power was calculated as the percentage of simulations with significant vaccine efficacy ($p < 0.05$) among scenarios with true vaccine efficacy. The values and parameters used to generate the unstimulated cell response and the vaccine effect were based on the magnitude of the cellular responses measured in the ANRS/INSERM VRI01 trial. All simulations were performed using SAS v. 9.4 (SAS Institute, Cary, NC).

2.3.Application to real data

We applied our modeling approach to analyze data from two HIV vaccine trials—VRI02 ANRS 149 LIGHT (NCT01492985) and ANRS/INSERM VRI01 (NCT02038842).

VRI02 ANRS 149 LIGHT evaluated a prime-boost combination of DNA-GTU and Lipopeptide vaccine followed by supervised treatment interruption (STI) in a therapeutic HIV phase II randomized trial. A total of 103 HIV-1-infected participants on c-ART were randomized (2:1 ratio) to receive three doses of DNA GTU-MultiHIV B (encoding Rev, Nef, Tat, Gag, and gp160) at week (W) 0, W4 and W12 followed by two doses of LIPO-5 vaccine containing long peptides from Gag, Pol, and Nef at W20 and W24, or a placebo. The HIV-specific CD4+ and CD8+ T-cell responses (IFNγ, IL2, TNFα) to HIV peptide pools (Gag, Nef, and Pol/Env) were assessed by ICS at W0 and W28 in 57 vaccinated and 32 placebo participants.

ANRS/INSERM VRI01 was a prophylactic open-label phase I/II randomized multicenter trial of the immunogenicity and safety of three candidate vaccines used as prime or boost: MVA HIV-B (encoding Gag, Pol, and Nef); LIPO-5 (five lipopeptides from Gag, Pol, and Nef); and DNA GTU-MultiHIV B (encoding Rev, Nef, Tat, Gag, and gp160 clade B). Healthy adult volunteers were randomized to four parallel groups: G1 received MVA at W0/8 + LIPO-5 at W20/28; G2, LIPO-5 at W0/8 + MVA at W20/28; G3, DNA at W0/4/12 + LIPO-5 at W20/28; and G4, DNA at W0/4/12 + MVA at W20/28. HIV-specific CD4+ and CD8+ T-cell responses (IFNγ, IL2, TNFα) were analyzed after stimulation of PBMC by HIV antigens (Gag, Nef, and Pol/Env peptide pools) using ICS at W0, W30 and W52. Only participants from G1, G2 and G4 were included (n = 62) by ICS because frequency of W30 IFNγ-ELISpot responders was 0% in G3. Details on the ICS are provided in Appendix C.

Model (1) for inter-group comparison (vaccine *vs*. placebo) at W28 (primary endpoint post-vaccination) and model (2) for intra-group comparison (W0–W28 in the vaccine group) were applied to VRI02 ANRS 149 LIGHT ICS data. For the ANRS/INSERM VRI01 trial, only model (2) was used for intra-group comparisons (W0–W30 [primary endpoint post-vaccination] in each vaccine group) as no formal comparison between vaccine regimens was

planned in this trial. Each model was run for each HIV-specific CD4+ and CD8+ T-cell response (IFNγ, IL2, and TNFα) and included the non-stimulated and the three stimulation conditions (Gag, Nef, and Pol/Env) as dependent variables (*i.e.*, a total of four response variables in the multivariate model).

3. Results

3.1. Simulation study

The statistical performance was similar for models (1) and (2) in terms of control of bias, type-I error, and statistical power. A summary of the statistical performance of models (1) and (2) is shown in Figure 2 and Appendix B – Figure 1, respectively, and detailed results for model (1) and model (2) are provided in Appendix B (Figures 2–7 for model (1) and Figures 8–13 for model (2)).

The type-I error rate was controlled at ≈ 5% in all scenarios by the model as well as the conventional approach using the raw stimulated cell response. For the conventional method with subtraction of the non-stimulated cell response, the type-I error rate was not controlled (> 20%) when there was a vaccine effect on the non-stimulated cell response (Figure 2C, upper panel). Regarding the control of bias and statistical power (lower panel), the performance of the two conventional approaches varied across the scenarios: the raw stimulated cell response performed better than the background-subtracted response in the absence of a correlation (Figure 2A) and, more importantly, in cases of a vaccine effect on the non-stimulated cell response (Figure 2C). In contrast, a conventional approach with background subtraction was better in cases of correlations between the non-stimulated and stimulated cell responses (Figure 2 B). The proposed modeling approach had excellent performance in all scenarios, at least as good as the respective conventional approach. The exception were cases of a vaccine effect on the non-stimulated response, in which the conventional approach without background-subtraction had slightly higher statistical power (Figure 2C).

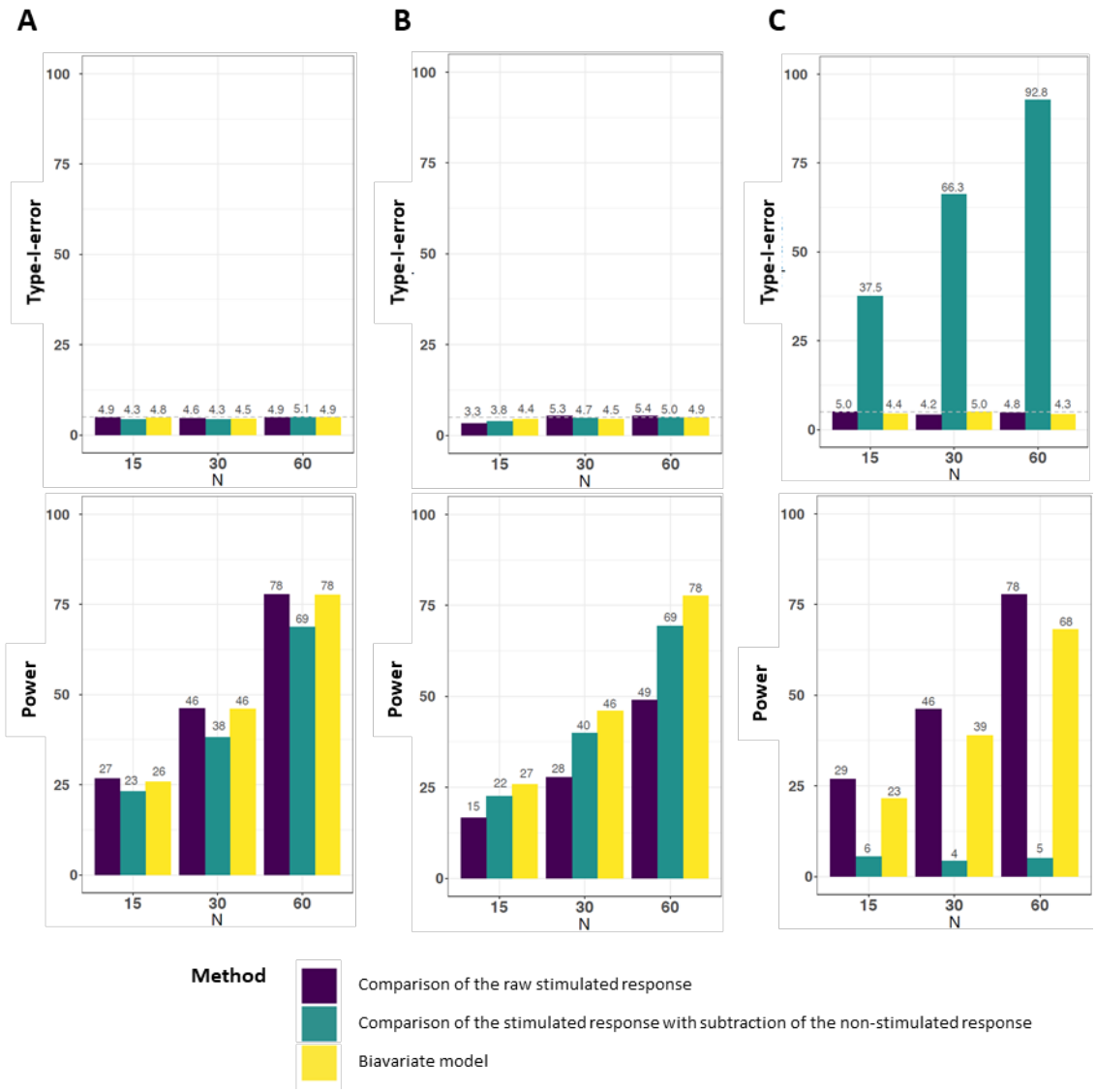


Figure 2: Evaluation of the performance of model 1 in terms of control of type 1 error and statistical power of the bivariate model for evaluating the effect of the vaccine (arm-vaccine versus placebo) compared to conventional approaches (with or without subtraction of the non-stimulated cell response) via simulations (1000 simulations per scenario).

To have good statistical performance, a model must control the type-I error at the nominal testing level (conventionally 5%) and the power must be as high as possible.
A: Scenarios with no correlation between stimulated and non-stimulated cell response and no vaccine effect on the non-stimulated cell response
B: Scenarios with correlation (p=0.70) between stimulated and non-stimulated cell response
C: Scenarios with a vaccine effect on the non-stimulated cell response

3.2. Application to real-world data

3.2.1. VRI02 ANRS 149 LIGHT

We applied the proposed model to evaluation of cellular immunogenicity at W28 for inter-arm comparison in the vaccine and placebo arms—model (1)—and intra-arm comparison in the vaccine and placebo arms—model (2). The estimated vaccine effect on cells stimulated by Gag, Nef, and Pol/Env using models 1 and 2 and the respective conventional approaches are presented in Figure 3.

Regarding the inter-arm comparison (Figure 3A), differences were observed between the two conventional approaches: a significant change in the frequency of CD4+ T cells producing TNFα, IL2, and IFNγ in response to the Pol/Env peptide pool was found in the vaccine arm *vs*. the placebo arm at W28 using the approach with subtraction of the non-stimulated cell response. In contrast, the approach without subtraction found only a significant change in the frequency of CD4+ T cells producing IFNγ in response to the same peptides. The bivariate modelling approach found specific CD4+ T cells producing IL2 in response to the Gag peptide pool. No significant change in CD8+ T-cell responses was found with the model or the conventional approaches with and without subtraction of the non-stimulated cell response. Regarding the intra-arm comparison in the vaccine arm (Figure 3B, upper panel), the results were similar for the three methods between W0 and W28—CD8+ T-cell responses to Nef and Pol/Env and CD4+ T-cell responses to the three peptides. A significant vaccine effect on CD8+ T cells producing IL2 was detected by bivariate modelling and the conventional approach with subtraction of the non-specific response but not with the non-subtracted conventional approach. In the placebo arm (Figure 3B, lower panel), the three methods yielded discordant results. Whereas no significant CD4+ and CD8+ T-cell responses were observed with the modeling approach, several significant cellular responses were found using the conventional approaches, particularly with subtraction of the unstimulated cell response. This result may be linked to an increased frequency of type-I errors with conventional approaches.

Estimates of the vaccine effect on the non-stimulated cell response and the association of the non-stimulated response with the stimulated cell responses are shown in Appendix D. No significant vaccine effect on the non-stimulated cell response was observed with model 1 (Appendix D, Figure 1) or model 2 (Appendix D, Figure 2). However, a significant association between the non-stimulated cell response and the stimulated cell responses was found for several CD4+ and CD8+ T-cell responses in both models. For example, estimates (standard deviation) from model 1 (inter-arm comparison at W28) for the analysis of the CD4+ IL2+ response after Pol/Env stimulation were 0.010 (0.033) for the vaccine effect on the stimulated cell responses ($p = 0.002$), and 1.097 (0.154) for the effect of the non-stimulated response on the stimulated response ($p < 0.001$).

Regarding conventional approaches without subtraction, the average difference (standard deviation) in raw stimulated responses between the two arms was 0.006 (0.018) ($p = 0.14$; standardized effect: 0.328). With the conventional approach with subtraction of the non-stimulated response, the average difference (standard deviation) in raw stimulated responses between the two arms was 0.010 (0.014) ($p = 0.003$; standardized effect: 0.684). For this example, the correlation coefficient between non-stimulated response and stimulated response was 0.57, which is a moderate correlation explaining why the conventional approach with subtraction of the non-stimulated cell response was closer of the modelling result.

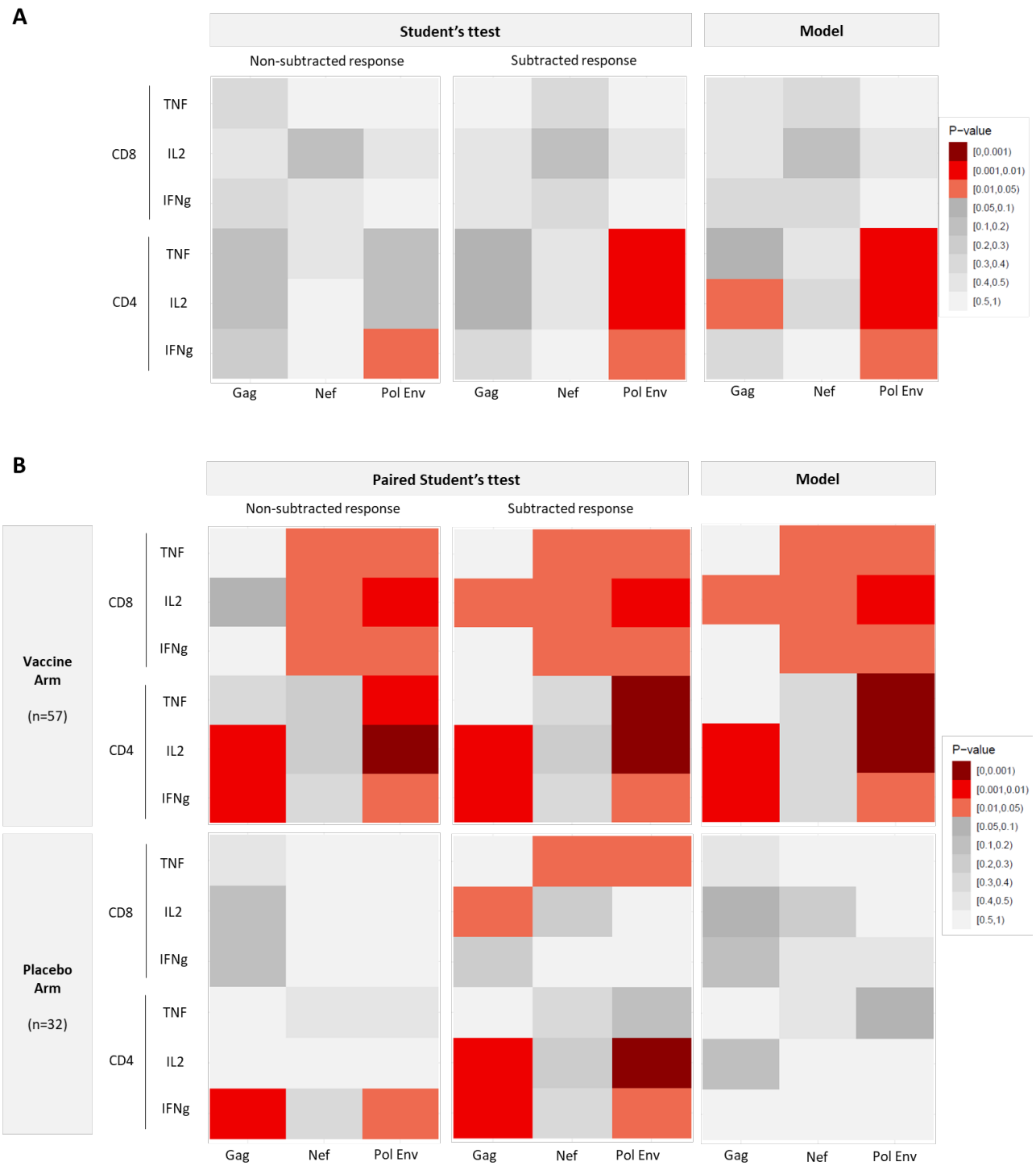


Figure 3. Heatmap of the p-values of the vaccine effect on the CD4+ and CD8+ T-cell responses measured by ICS in the VRI02 ANRS 149 LIGHT trial arms. The three approaches were (1) comparison of the raw stimulated cell response, (2) comparison of the specific response with subtraction of the non-stimulated cell response, and (3) a bivariate model with the non-stimulated cell responses and stimulated cell responses as dependent variables.

A. Inter-arm (vaccine *vs*. placebo) comparisons.

B. Intra-arm (week 28 *vs*. baseline) comparisons in the vaccine and placebo arms.

3.2.2. ANRS/INSERM VRI01

We applied the proposed model to evaluate cellular immunogenicity at W30 in the MVA-LIPO, LIPO-MVA, and GTU-MVA arms. Estimates of the vaccine effects on the Gag, Nef, and Pol/Env responses in each arm (model 2) and the respective conventional approaches are shown in Figure 4.

As expected, some discrepancies were observed between the two conventional approaches for measuring the vaccine effect. The modelling approach resolved this uncertainty, and yielded more robust results that were sometimes closer to one conventional approach, sometimes closer to the other.

Estimates of the vaccine effect on the non-stimulated response and the effect of the non-stimulated response on the stimulated responses are presented in Appendix E – Figure 1. In the analysis of VRI02 ANRS 149 LIGHT, no significant vaccine effect on non-stimulated cell responses was observed. A significant association of the non-stimulated response with the stimulated responses was found for several CD4+ and CD8+ T-cell responses.

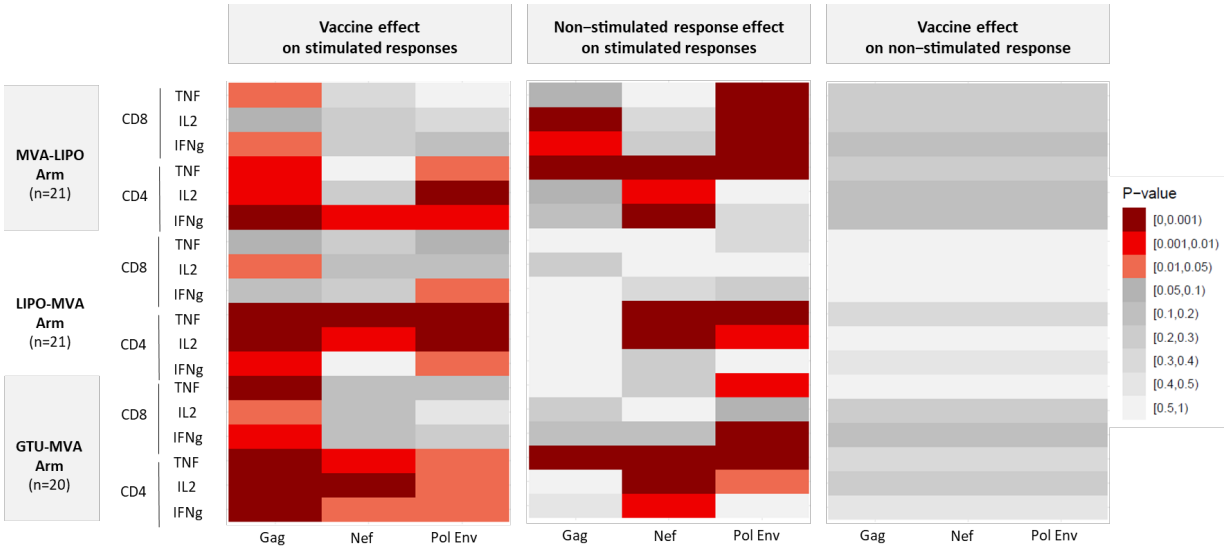


Figure 4. Heatmap of the *p*-values of the intra-arm analysis of the vaccine effect on the CD4+ and CD8+ T-cell responses measured by ICS in the ANRS/INSERM VRI01 trial arms. The three approaches were (1) comparison of the raw stimulated response between

W30 and baseline, (2) comparison of the stimulated response with subtraction of the non-stimulated response between W30 and baseline, and (3) a bivariate model with the non-stimulated responses and the stimulated responses as dependent variables.

4. Discussion

We present a multivariate modelling approach to analyze the cellular immune response of vaccine candidates during vaccine clinical trials. Simulations showed that the bivariate model effectively controlled the type-I error in all assessed scenarios with different population sizes, while its statistical power was at least as good as the conventional approaches in all scenarios. Controlling type-I errors is fundamental for the consistency of research. A high frequency of type-I errors leads to an apparently statistically significant result that is not reproducible in further studies. This is one of the determinants of the very high prevalence of false-positive results (20). The lack of statistical power is also a drawback, especially in the context of early phase vaccine clinical trials in which the number of subjects is restricted. Conventional approaches based on comparison of the background-subtracted response by t-test are the most frequently used (1-4) but did not control type-I error and had low statistical power for the vaccine effect on the non-stimulated cell response. Comparison of the data without subtraction of the non-stimulated response exhibited little statistical power in cases of correlations between the non-stimulated and stimulated cell responses.

The use of real data from prophylactic and therapeutic HIV vaccine trials showed the feasibility of the modelling approach. Not surprisingly, divergent results among the three analysis approaches (modelling and two conventional approaches) were obtained for some ICS responses. This demonstrated that the proposed model yields robust results and provides information on the correlation between the non-stimulated and stimulated cell responses. The validity of the proposed modelling approach could not be tested with real data but an *in silico* study demonstrated the drawbacks of conventional methods (Gilbert *et al.*). The systematic use

of one of the conventional approaches leads to erroneous results. The originality of our approach lies in the simultaneous modeling of the non-stimulated and stimulated responses, unlike conventional approaches. This enables assessment of the vaccine effect on stimulated cell response adjusted for the non-stimulated cell response as well as of the vaccine effect on the non-stimulated cell response without loss of statistical power. The strength of the model compared to conventional approaches is its good statistical performance irrespective of the relationship between the vaccine and the non-stimulated response and between the non-stimulated response and the stimulated response. The conventional approaches do not take these relationships into account. In addition, the bivariate model is easily extendable to multivariate models with more than two dependent variables, allowing consideration of more than one antigen, while a large number of tests must be performed using conventional approaches, leading to a risk of type-I error. Furthermore, while controlling for statistical errors, the model provides more biological information on the effect of variables on the background or on the specific response independently of the intervention. Notably, the same modelling approach can be used to explore any variable that modifies the effect of interest (e.g., the vaccine) on the ICS response. As an example, it could be used to evaluate whether the vaccine induced a similar ICS response in women and men through an interaction term.

One potential drawback of our approach is the fact that it is a fully parametric method making assumptions. It is assuming that residuals are normally (i.e., Gaussian) distributed with constant variance, which could not be the case. However, there are several arguments for thinking that the consequences on the usefulness of the method are limited. First, the conditional distribution could be more likely Gaussian than the marginal one tested with the t-test. Second, the mixed models are quite robust to misspecification of residual distribution (21). Therefore, we think that parametric assumptions should not limit the use of the approach if the sample size is adequate (< 20 individuals).

Although our modeling approach is more complex than conventional approaches based on basic statistical tests (Student's *t*-test), this should not prevent its practical use. To facilitate its implementation, the code for SAS and R software is provided in the Appendix. In addition, we developed an R-shiny application with an interactive and user-friendly web interface that enables immunologists to analyze their data with no statistical software or experience required. The application provides simple and clear interpretation of the results and the output can be directly integrated into publications.

In conclusion, our novel method of analyzing T-cell immunogenicity data, based on bivariate modelling, enables consideration of all available information with more flexibility than conventional methods, leading to more accurate and more detailed results, enhancing interpretation of the vaccine effect on T-cell function. The use of conventional approaches, in particular comparison of the T-cell response after background subtraction, should no longer be recommended for ICS in vaccine trials. Our multivariate model is therefore an alternative to conventional approaches. The bivariate model could be used to analyze any type of functional response in which a non-stimulated cell response is measured.

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see:

http://www.textcheck.com/certificate/Sq7OxQ

Acknowledgments

The authors have no competing interests to declare.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. For this research, the SISTM team receives funding from the The Vaccine Research Institute (VRI), Laboratory of excellence supported by the Investissements d'Avenir program managed by the ANR under reference ANR-10-LABX-77-01.

Declarations of interest: none

References

1. Kibuuka H, Berkowitz NM, Millard M, Enama ME, Tindikahwa A, Sekiziyivu AB, et al. Safety and immunogenicity of Ebola virus and Marburg virus glycoprotein DNA vaccines assessed separately and concomitantly in healthy Ugandan adults: a phase 1b, randomised, double-blind, placebo-controlled clinical trial. The Lancet 2015;385(9977):1545-54.
2. Guardo A, Gómez CE, Díaz-Brito V, Pich J, Arnaiz JA, Perdiguero B, et al. Safety and vaccine-induced HIV-1 immune responses in healthy volunteers following a late MVA-B boost 4 years after the last immunization. PLoS One 2017;12(10).
3. Lévy Y, Thiébaut R, Montes M, Lacabaratz C, Sloan L, King B, et al. Dendritic cell-based therapeutic vaccine elicits polyfunctional HIV-specific T-cell immunity associated with control of viral load. Eur J Immunol 2014;44(9):2802-10.
4. Luabeya AKK, Kagina BMN, Tameris MD, Geldenhuys H, Hoff ST, Shi Z, et al. First-in-human trial of the post-exposure tuberculosis vaccine H56:IC31 in Mycobacterium tuberculosis infected and non-infected healthy adults. Vaccine 2015;33(33):4130-40.
5. Horton H, Thomas E, Stucky J, Frank I, Moodie Z, Huang Y, et al. Optimization and Validation of an 8-Color Intracellular Cytokine Staining (ICS) Assay to Quantify Antigen-Specific T Cells Induced by Vaccination. J Immunol Methods 2007;323(1):39-54.
6. De Rosa SC. Vaccine applications of flow cytometry. Methods 2012;57(3):383-91.
7. Lin L, Finak G, Ushey K, Seshadri C, Hawn TR, Frahm N, et al. COMPASS identifies T-cell subsets correlated with clinical outcomes. Nat Biotechnol 2015;33(6):610-6.
8. Churchyard GJ, Morgan C, Adams E, Hural J, Graham BS, Moodie Z, et al. A phase IIA randomized clinical trial of a multiclade HIV-1 DNA prime followed by a multiclade rAd5 HIV-1 vaccine boost in healthy adults (HVTN204). PLoS ONE 2011;6(8):e21225.
9. Tung FY, Tung JK, Pallikkuth S, Pahwa S, Fischl MA. A therapeutic HIV-1 vaccine enhances anti-HIV-1 immune responses in patients under highly active antiretroviral therapy. Vaccine 2016;34(19):2225-32.
10. Janes HE, Cohen KW, Frahm N, De Rosa SC, Sanchez B, Hural J, et al. Higher T-Cell

Responses Induced by DNA/rAd5 HIV-1 Preventive Vaccine Are Associated With Lower HIV-1 Infection Risk in an Efficacy Trial. J Infect Dis 2017;215(9):1376-85.
11. Kalams SA, Parker SD, Elizaga M, Metch B, Edupuganti S, Hural J, et al. Safety and Comparative Immunogenicity of an HIV-1 DNA Vaccine in Combination with Plasmid Interleukin 12 and Impact of Intramuscular Electroporation for Delivery. J Infect Dis 2013;208(5):818-29.
12. Richert L, Thiebaut R, BiostatisticsBioinformatics Division of the ANRS Vaccine Research Institute. Is the current use of « positivity » thresholds meaningful for evaluating HIV-vaccine immunogenicity endpoints? AIDS 2013;27(8):1362-5.
13. Liu AY, De Rosa SC, Guthrie BL, Choi RY, Kerubo-Bosire R, Richardson BA, Kiarie J, Farquhar C, Lohman-Payne B High background in ELISpot assays is associated with elevated levels of immune activation in HIV-1-seronegative individuals in Nairobi. Immun Inflamm Dis 2018;6(3):392-401.
14. Su LF, Kidd BA, Han A, Kotzin JJ, Davis MM. Virus-specific CD4+ memory phenotype T cells are abundant in unexposed adults. Immunity 2013;38(2):373-83.
15. Gilbert PB, Sato A, Sun X, Mehrotra DV. Efficient and robust method for comparing the immunogenicity of candidate vaccines in randomized clinical trials. Vaccine 2009;27(3):396-401.
16. Thiébaut R, Jacqmin-Gadda H, Chêne G, Leport C, Commenges D. Bivariate linear mixed models using SAS proc MIXED. Computer Methods and Programs in Biomedicine 2002;69(3):249-56.
17. Moodie Z, Rossini AJ, Hudgens MG, Gilbert PB, Self SG, Russell ND. Statistical evaluation of HIV vaccines in early clinical trials. Contemporary Clinical Trials 2006;27(2):147-60.
18. Richert L, Lhomme E, Fagard C, Lévy Y, Chêne G, Thiébaut R. Recent developments in clinical trial designs for HIV vaccine research. Hum Vaccin Immunother. 9 mars 2015;11(4):1022-9.
19. VICI: acurate estimation of Vaccine Induced Cellular Immunogenicity by bivariate modelling [Internet]. https://shiny-vici.apps.math.cnrs.fr/
20. Ioannidis JP. Why most discovered true associations are inflated. Epidemiology. 2008;19(5):640-8.
21. Jacqmin-Gadda H, Sibillot S, Proust C, Molina J-M, Thiébaut R. Robustness of the linear mixed model to misspecified error distribution. Computational Statistics & Data Analysis 2007;51(10):5142-54.

Appendix A: Model specification and software codes

SAS code

Model (1)

Below an example of SAS code for a bivariate model for an inter-arm comparison at one time point post vaccination where STIMULATION is a binary variable indicating the response variable concerned (STIMULATION = 0 for Y_i^{NS} and STIMULATION = 1 for Y_i^{S}); ID_PATIENT a single identification number for each patient, VACCINE a qualitative variable for the treatment group; Y_NS with Y_NS = 0 when STIMULATION = 0 and the value of the non-stimulated response when STIMULATION = 1.

```
proc mixed data = DATATABLE;
class ID_PATIENT STIMULATION ;
model Y = STIMULATION STIMULATION*VACCINE STIMULATION* Y_NS / cl noint ;
repeated /type=VC grp=STIMULATION sub=ID_PATIENT ;
run ;
```

Model (2)

Below an example of SAS code for a bivariate model for an intra-arm comparison (post vaccination versus baseline) where STIMULATION is a binary variable indicating the response variable concerned (STIMULATION = 0 for Y_i^{NS} and STIMULATION = 1 for Y_i^{S}); ID_PATIENT a single identification number for each patien; Y_d_NS with Y_d_NS = 0 when STIMULATION = 0 and the value of the differential (post-vaccination / baseline) of the non-stimulated response when STIMULATION = 1.

```
proc mixed data = DATATABLE;
class ID_PATIENT STIMULATION ;
model Y_d  = STIMULATION STIMULATION*Y_d_NS/ cl noint;
repeated /type=VC grp= STIMULATION sub= ID_PATIENT ;
run ;
```

R code

Below an example of R code for a bivariate model for an inter-arm comparison at one time point post vaccination where STIMULATION is a binary variable indicating the response variable concerned (STIMULATION = 0 for Y_i^{NS} and STIMULATION = 1 for Y_i^{S}); ID_PATIENT a single identification number for each patient, VACCINE a qualitative variable for the treatment group; Y_NS with Y_NS = 0 when STIMULATION = 0 and the value of the non-stimulated response when STIMULATION = 1.

Model (1)

```
nlme::gls(Y ~ -1 + STIMULATION*ARM + Y_NS,
                    data = DATATABLE,
                    weights = nlme::varIdent(value = c("1" = 1), form = ~ 1 | STIMULATION),
                    method="REML"
        )
```

Model (2)

```
nlme::gls(Y_d ~ -1 + STIMULATION + Y_d_NS,
                    data = DATATABLE,
                    weights = nlme::varIdent(value = c("1" = 1), form = ~ 1 | STIMULATION),
```

```
                method="REML"
    )
```

Data set

Example of dataset for running model (1)

USUBJID	Y	ARM	Y_NS	STIMULATION
1	0.00339	0	0	0
1	0.0569	0	0.00339	1
2	0.0093	0	0	0
2	0.285	0	0.0093	1
3	0.0128	1	0	0
3	0.0517	1	0.0128	1
4	0.0118	1	0	0
4	0.135	1	0.0118	1
5	0.00444	1	0	0
5	0.27	1	0.00444	1
6	0.00488	1	0	0
6	0.0399	1	0.00488	1
7	0.00229	1	0	0
7	0.0743	1	0.00229	1
8	0.00256	0	0	0
8	0.0282	0	0.00256	1
9	0.00481	0	0	0
9	0.119	0	0.00481	1
10	0.00662	0	0	0
10	0.0416	0	0.00662	1

Appendix B – Simulations

Scenarios

Appendix B – Table 1. Description of the different scenarios of simulation for Model 1 and Model 2 and parameters used for each scenario

Scenarios of simulation	**Hypothesis**	Y_i^{NS}			Y_i^S			
		β_0^{NS}	β_1^{NS}	ε_i^{NS}	β_0^S	β_1^S	β_2^S	ε_i^S
Model (1), two arms comparison at one time point								
Similar vaccine effect between arms on stimulated response	Vaccine effect on stimulated response: H0 ($\beta_1^S = 0$) Vaccine effect on non-stimulated response: H0 ($\beta_1^{NS} = 0$) Non-stimulated response effect on stimulated response: H0 ($\beta_2^S = 0$)	0.02	0	$\sim N(0, 0.01)$	0.04	0	0	$\sim N(0, 0.02)$
Vaccine effect on stimulated response, no correlation between	Vaccine effect on stimulated response: H1 ($\beta_1^S \neq 0$)	0.02	0	$\sim N(0, 0.01)$	0.04	0	0	$\sim N(0, 0.02)$

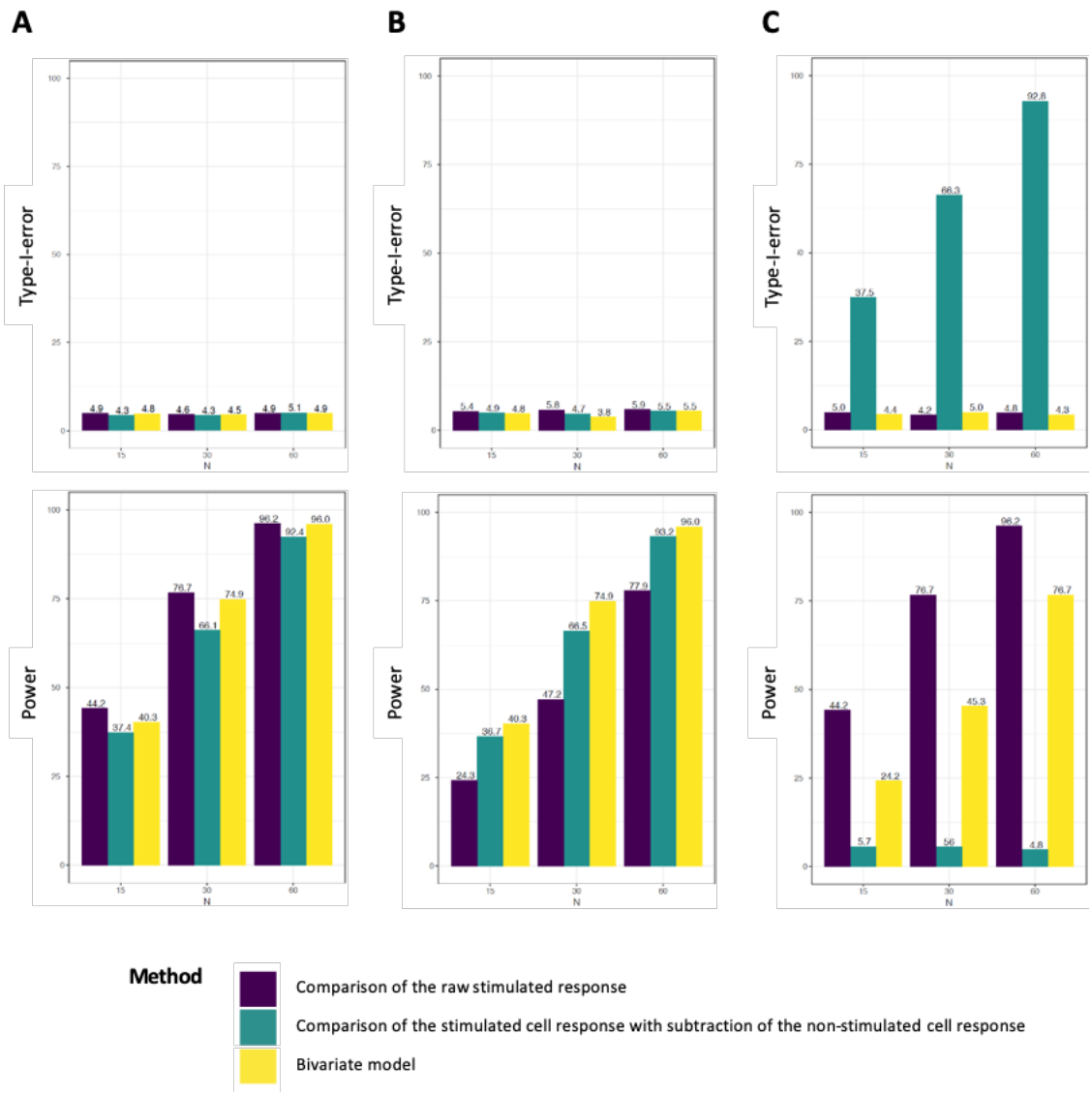
stimulated and non-stimulated response	Vaccine effect on non-stimulated response: H0 ($\beta_1^{NS} = 0$) Non-stimulated response effect on stimulated response: H0 ($\beta_2^{S} = 0$)							
Similar vaccine effect between arms on stimulated response, correlation between stimulated and non-stimulated response	Vaccine effect on stimulated response: H0 ($\beta_1^{S} = 0$) Vaccine effect on non-stimulated response: H0 ($\beta_1^{NS} = 0$) Non-stimulated response effect on stimulated response: H1 ($\beta_2^{S} \neq 0$)	0.02	0	$\sim N(0, 0.01)$	0.04	0	0	$\sim N(0, 0.02)$
Vaccine effect on stimulated response, no correlation between stimulated and non-stimulated response, correlation between stimulated and non-stimulated response	Vaccine effect on stimulated response: H1($\beta_1^{S} \neq 0$) Vaccine effect on non-stimulated response: H0 ($\beta_1^{NS} = 0$) Non-stimulated response effect on stimulated response: H1 ($\beta_2^{S} \neq 0$)	0.02	0	$\sim N(0, 0.01)$	0.04	0	0	$\sim N(0, 0.02)$

Similar vaccine effect between arms on stimulated response, Vaccine effect on non-stimulated response	Vaccine effect on stimulated response: H0 ($\beta_1^S = 0$) Vaccine effect on non-stimulated response: H1 ($\beta_1^{NS} \neq 0$) Non-stimulated response effect on stimulated response: H0 ($\beta_2^S = 0$)	0.02	0.01	$\sim N(0, 0.01)$	0.04	0.01	0.01	$\sim N(0, 0.02)$
Vaccine effect on stimulated response, Vaccine effect on non-stimulated response	Vaccine effect on stimulated response: H1 ($\beta_1^S \neq 0$) Vaccine effect on non-stimulated response: H1 ($\beta_1^{NS} \neq 0$) Non-stimulated response effect on stimulated response: H0 ($\beta_2^S = 0$)	0.02	0.01	$\sim N(0, 0.01)$	0.04	0.01	0.01	$\sim N(0, 0.02)$
Model (2), one arm comparison (post-vaccination versus baseline)								
No vaccine effect	Vaccine effect on stimulated response: H0 ($\beta_0^S = 0$)	0		$\sim N(0, 0.01)$	0	0		$\sim N(0, 0.02)$

	Vaccine effect on non-stimulated response: H0 ($\beta_0^{NS} = 0$) Non-stimulated response effect on stimulated response: H0 ($\beta_1^{S} = 0$)							
Vaccine effect on stimulated response, no correlation between stimulated and non-stimulated response	Vaccine effect on stimulated response: H1($\beta_0^{S} \neq 0$) Vaccine effect on non-stimulated response: H0 ($\beta_0^{NS} = 0$) Non-stimulated response effect on stimulated response: H0 ($\beta_1^{S} = 0$)	0		$\sim N(0, 0.01)$	0	0		$\sim N(0, 0.02)$
No vaccine effect between arms on stimulated response, correlation between stimulated and non-stimulated response	Vaccine effect on stimulated response: H0 ($\beta_0^{S} \neq 0$) Vaccine effect on non-stimulated response: H0 ($\beta_0^{NS} = 0$) Non-stimulated response effect on stimulated response: H1 ($\beta_1^{S} \neq 0$)	0		$\sim N(0, 0.01)$	0	0		$\sim N(0, 0.02)$

Vaccine effect on stimulated response, no correlation between stimulated and non-stimulated response, correlation between stimulated and non-stimulated response	Vaccine effect on stimulated response: H1 ($\beta_0^S \neq 0$) Vaccine effect on non-stimulated response: H0 ($\beta_0^{NS} \neq 0$) Non-stimulated response effect on stimulated response: H1 ($\beta_1^S \neq 0$)	0		$\sim N(0, 0.01)$	0	0		$\sim N(0, 0.02)$
Similar vaccine effect between arms on stimulated response, Vaccine effect on non-stimulated response	Vaccine effect on stimulated response: H0 ($\beta_0^S = 0$) Vaccine effect on non-stimulated response: H1 ($\beta_0^{NS} \neq 0$) Non-stimulated response effect on stimulated response: H0 ($\beta_1^S = 0$)	0.01		$\sim N(0, 0.01)$	0.01	0.01		$\sim N(0, 0.02)$
Vaccine effect on stimulated response, Vaccine effect on non-stimulated response	Vaccine effect on stimulated response: H1 ($\beta_0^S \neq 0$) Vaccine effect on non-stimulated response: H1 ($\beta_0^{NS} \neq 0$)	0.01		$\sim N(0, 0.01)$	0.01	0.01		$\sim N(0, 0.02)$

Non-stimulated response effect on stimulated response: H0 ($\beta_1^S \neq 0$)			



Appendix B – Figure 1: Evaluation of the performance of Model 2 in terms of control of type 1 error and statistical power of the bivariate model for evaluating the effect of the vaccine (arm-vaccine versus placebo) compared to conventional approaches (with or without subtraction of the non-stimulated response) via simulations (1000 simulations per scenario).

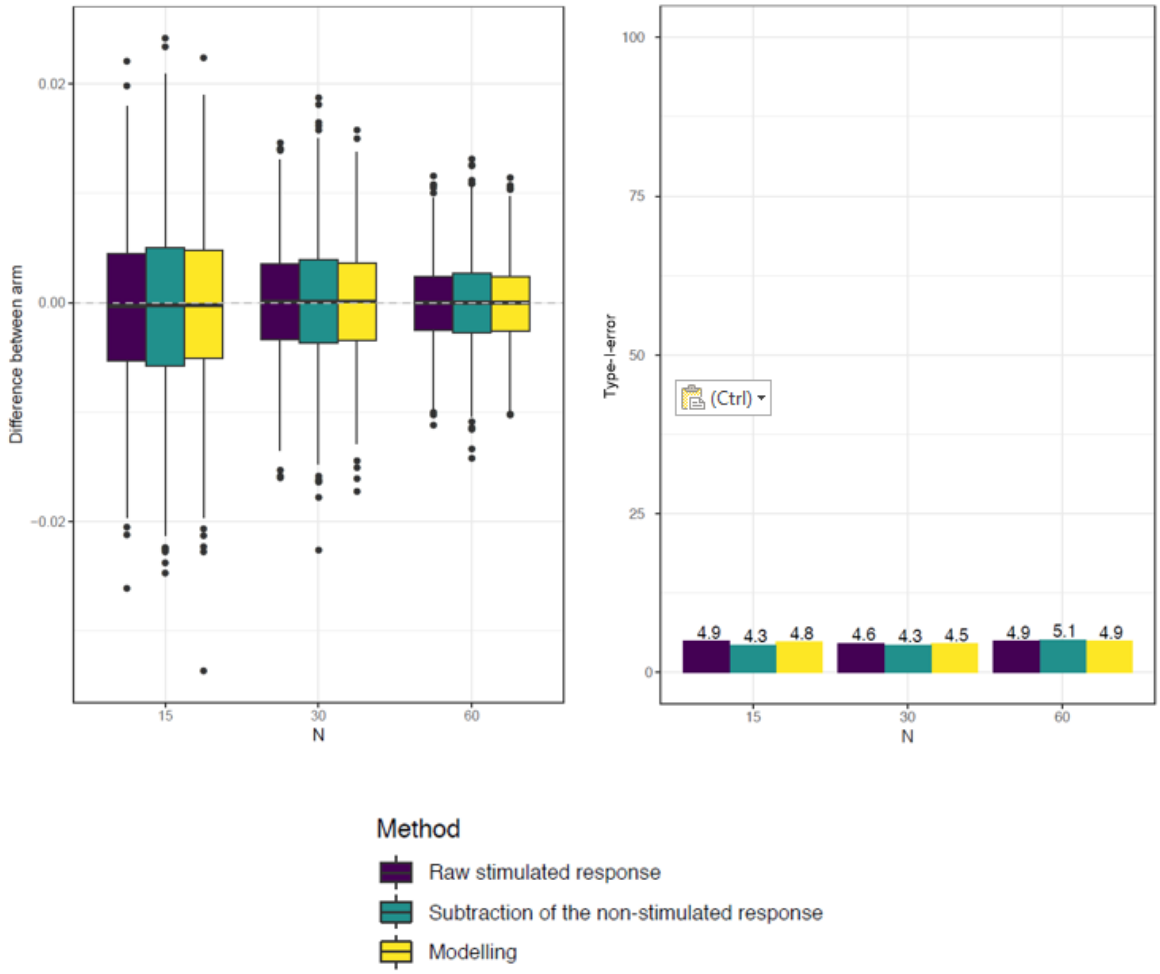
To have good statistical performance, a model must control the type 1 error at 5% and the power must be as high as possible (generally 80%).

A: Scenarios with no correlation between stimulated and non-stimulated response and no vaccine effect on the non-stimulated response

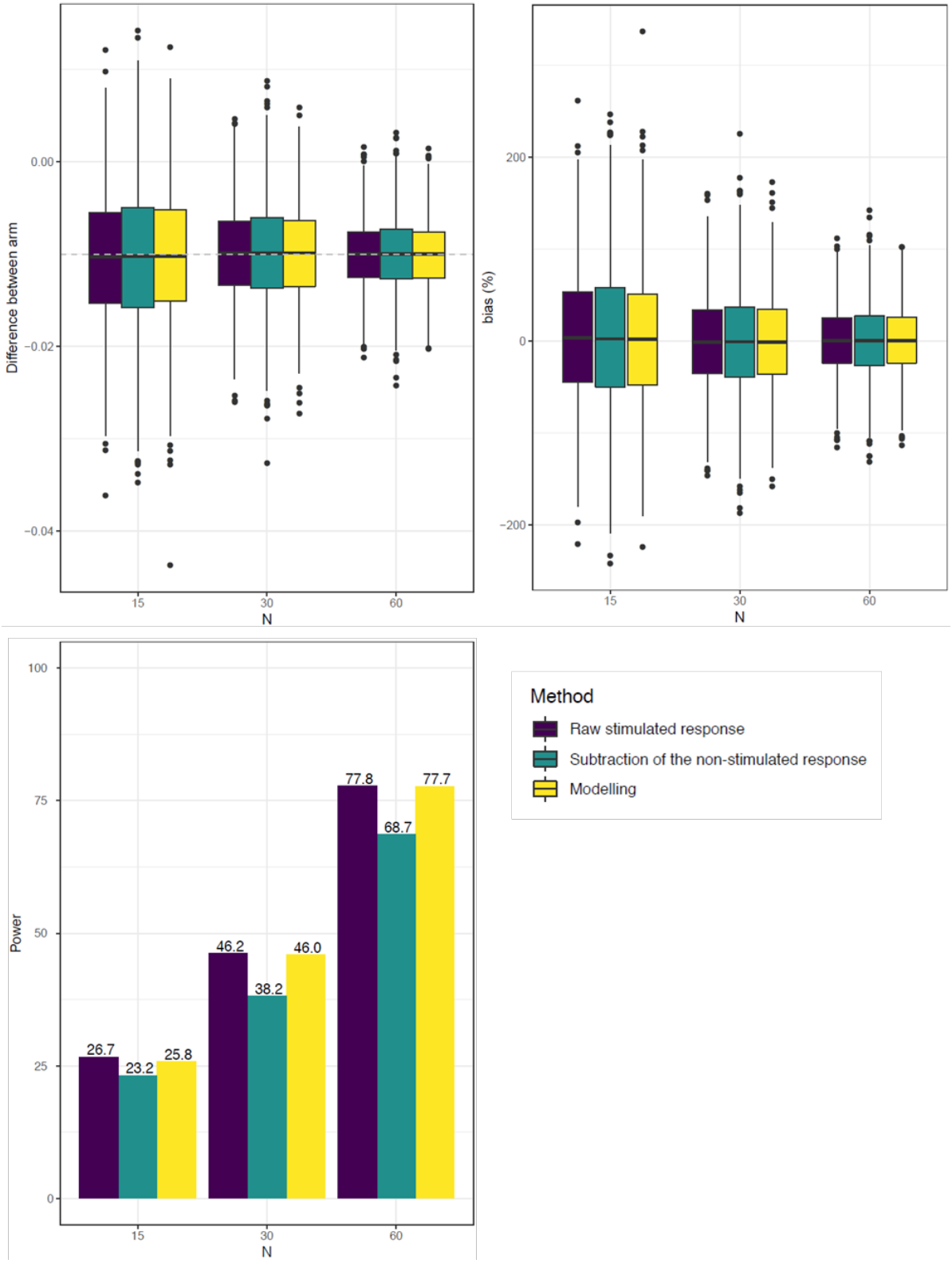
B: Scenarios with correlation (p=0.70) between stimulated and non-stimulated response

C: Scenarios with vaccine effect on the non-stimulated response

Simulation results – Model 1

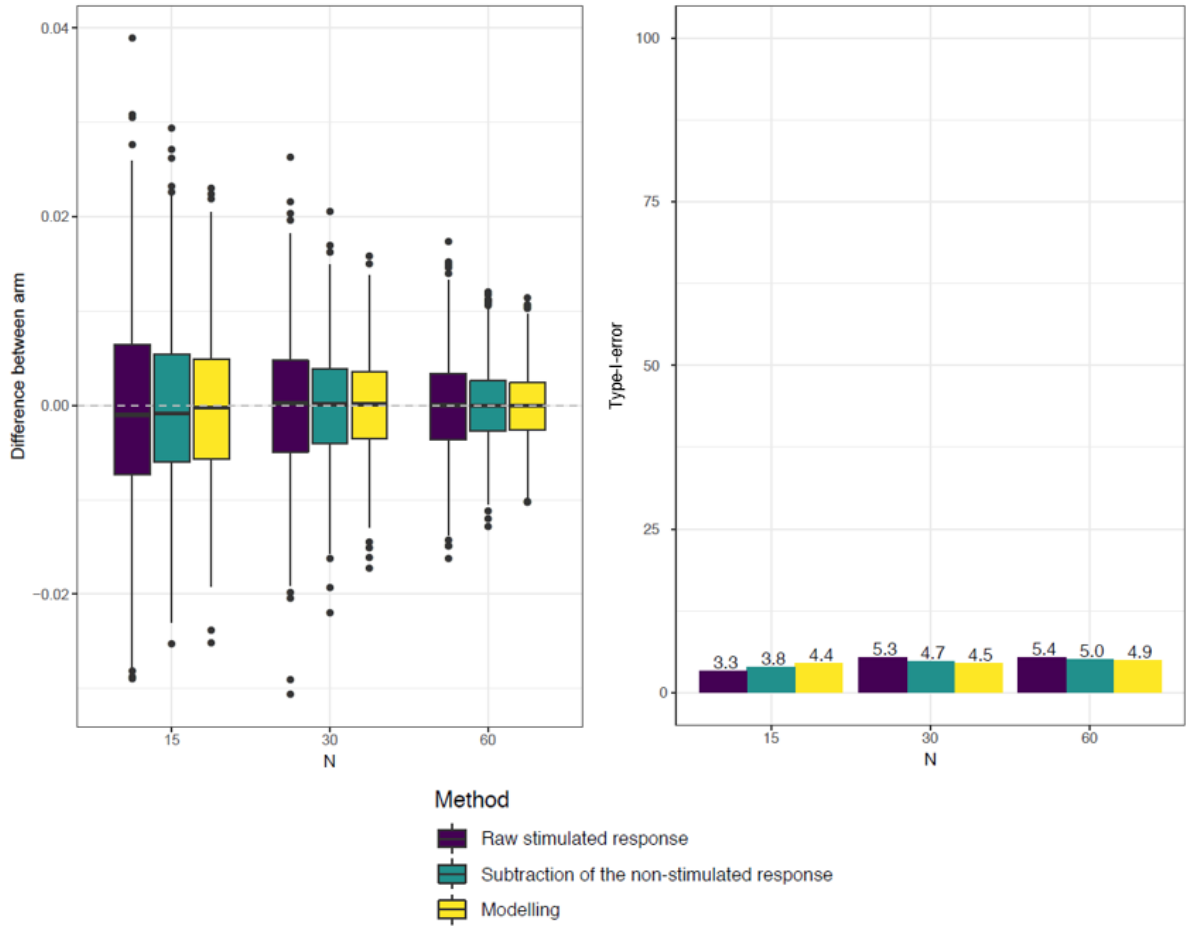


Appendix B – Figure 1. Scenario simulating a similar vaccine effect between arms on stimulated response. Comparison of the difference between arm (left) and the type I error risk (right) between the modelling approach (model 2) and two conventional approaches (comparing respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)

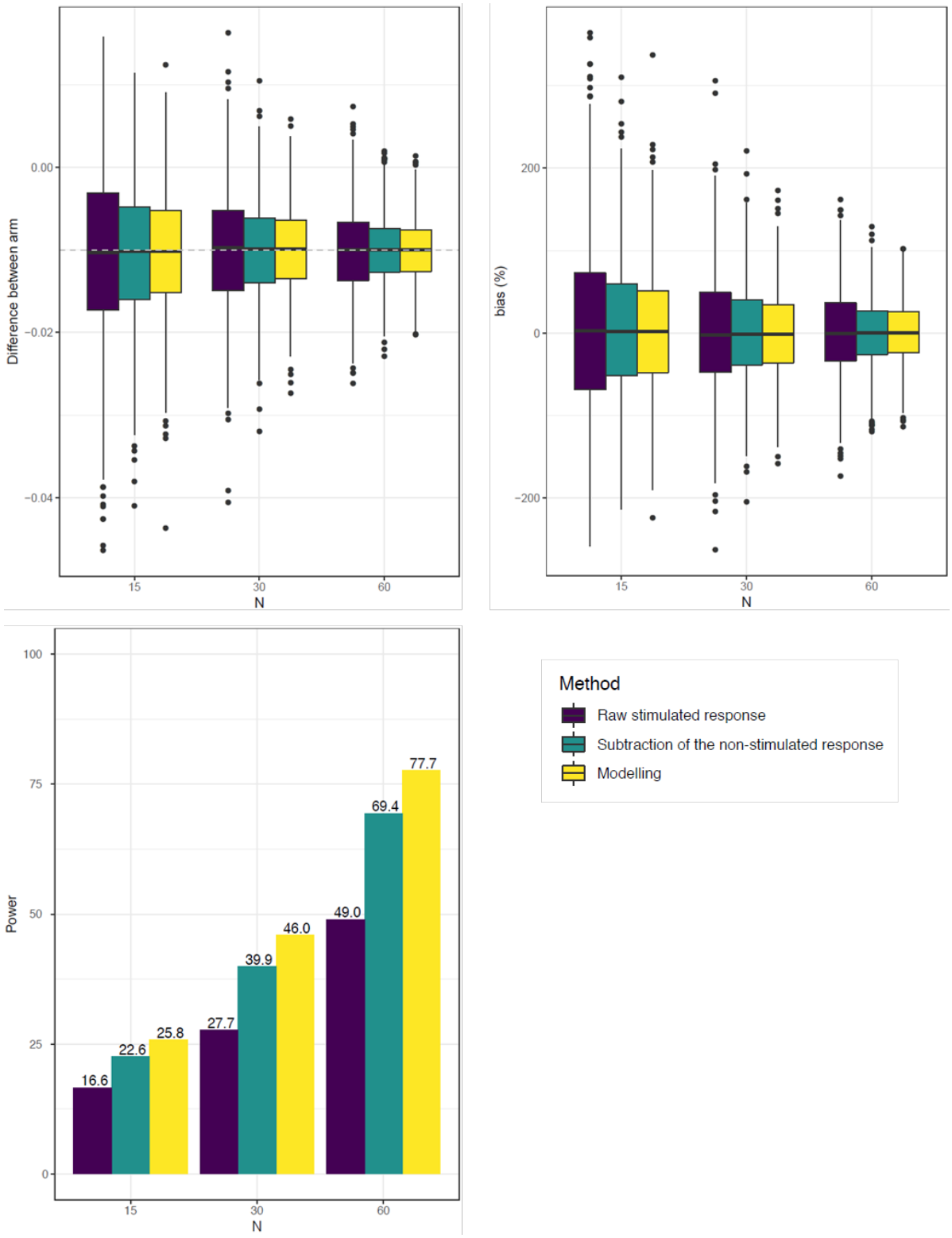


Appendix B – Figure 2. Scenario simulating a vaccine effect on stimulated response, no correlation between stimulated and non-stimulated response. Comparison of the difference between arm, the percentage of bias and the power between the modelling approach (model 2) and two conventional

approaches (comparing respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)

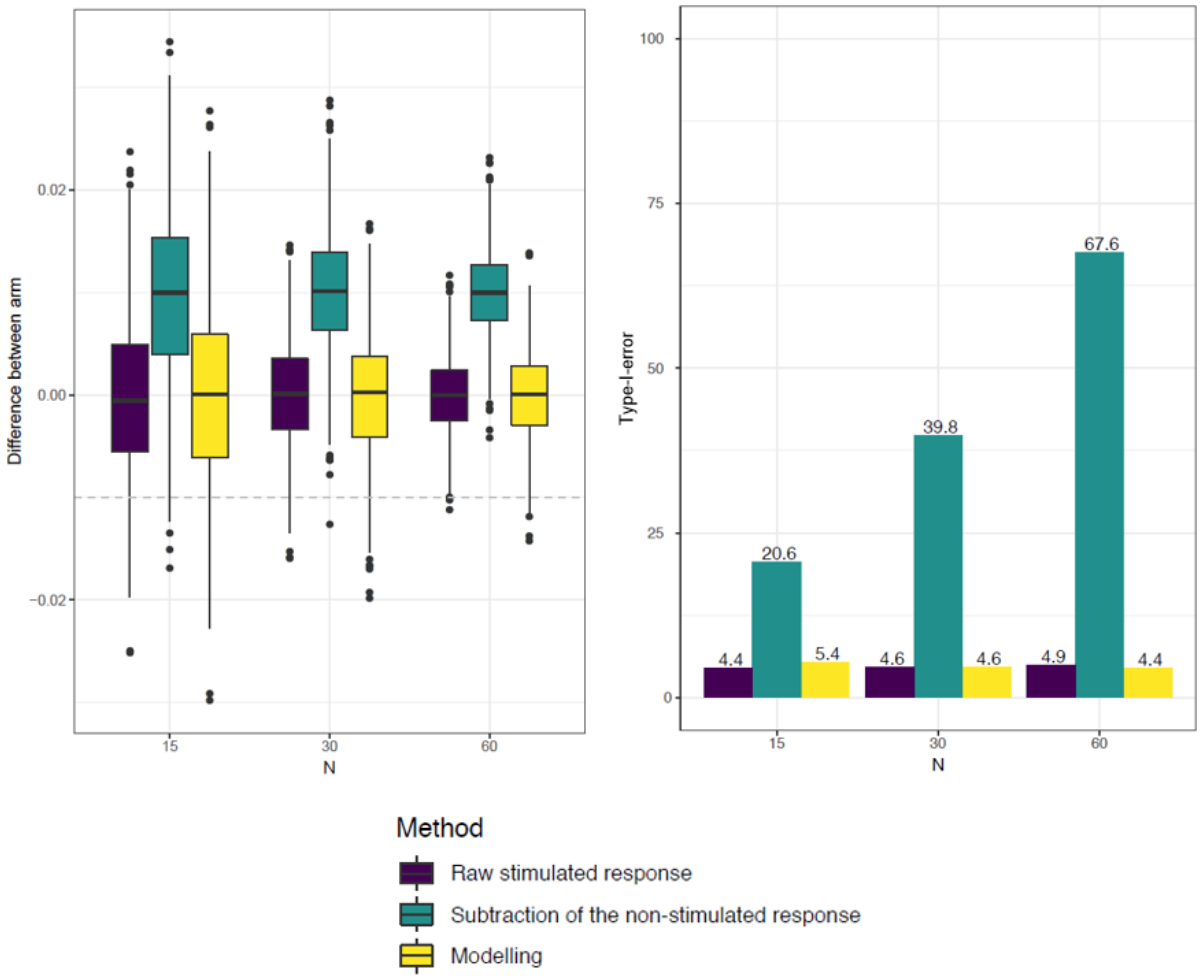


Appendix B – Figure 3. Scenario simulating a similar vaccine effect between arms on stimulated response, correlation between stimulated and non-stimulated response. Comparison of the difference between arm (left) and the type I error risk (right) between the modelling approach (model 2) and two conventional approaches (comparing respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)

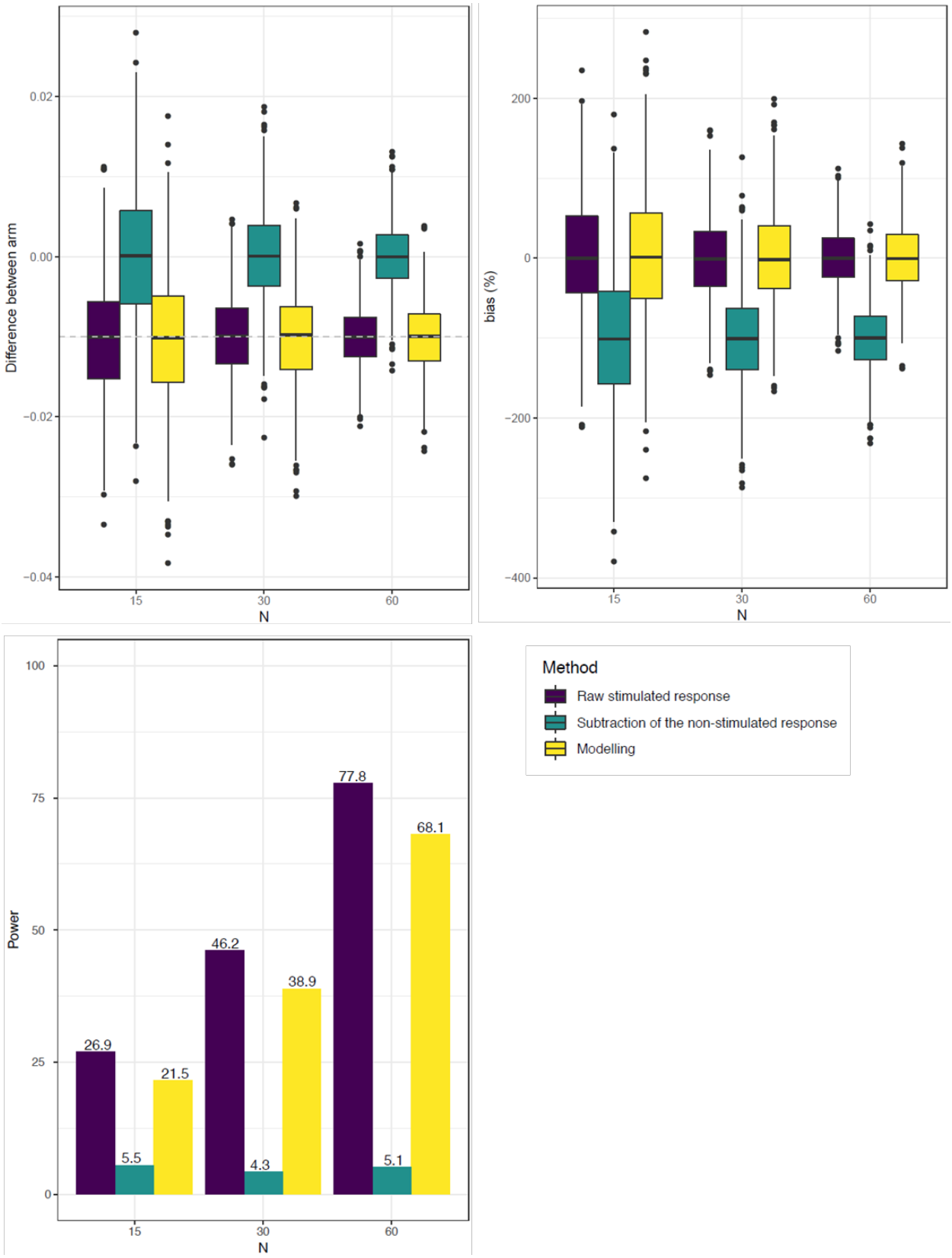


Appendix B – Figure 4. Scenario simulating a vaccine effect on stimulated response, no correlation between stimulated and non-stimulated response, correlation between stimulated and non-stimulated response. Comparison of the difference between arm, the percentage of bias and the power between the

modelling approach (model 2) and two conventional approaches (comparing respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)



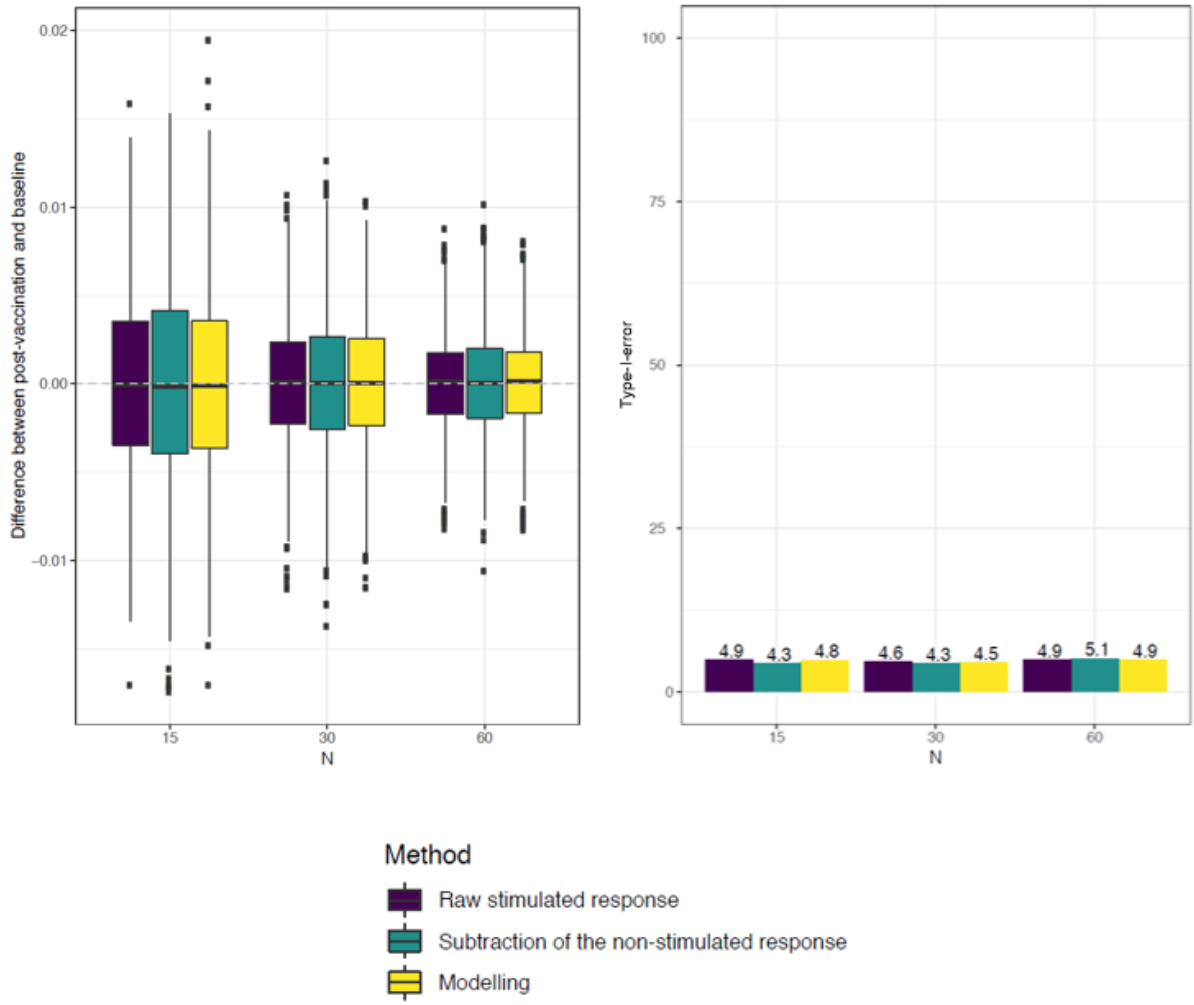
Appendix B – Figure 5. Scenario simulating a similar vaccine effect between arms on stimulated response, Vaccine effect on non-stimulated response. Comparison of the difference between arm (left) and the type I error risk (right) between the modelling approach (model 2) and two conventional approaches (comparing respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)



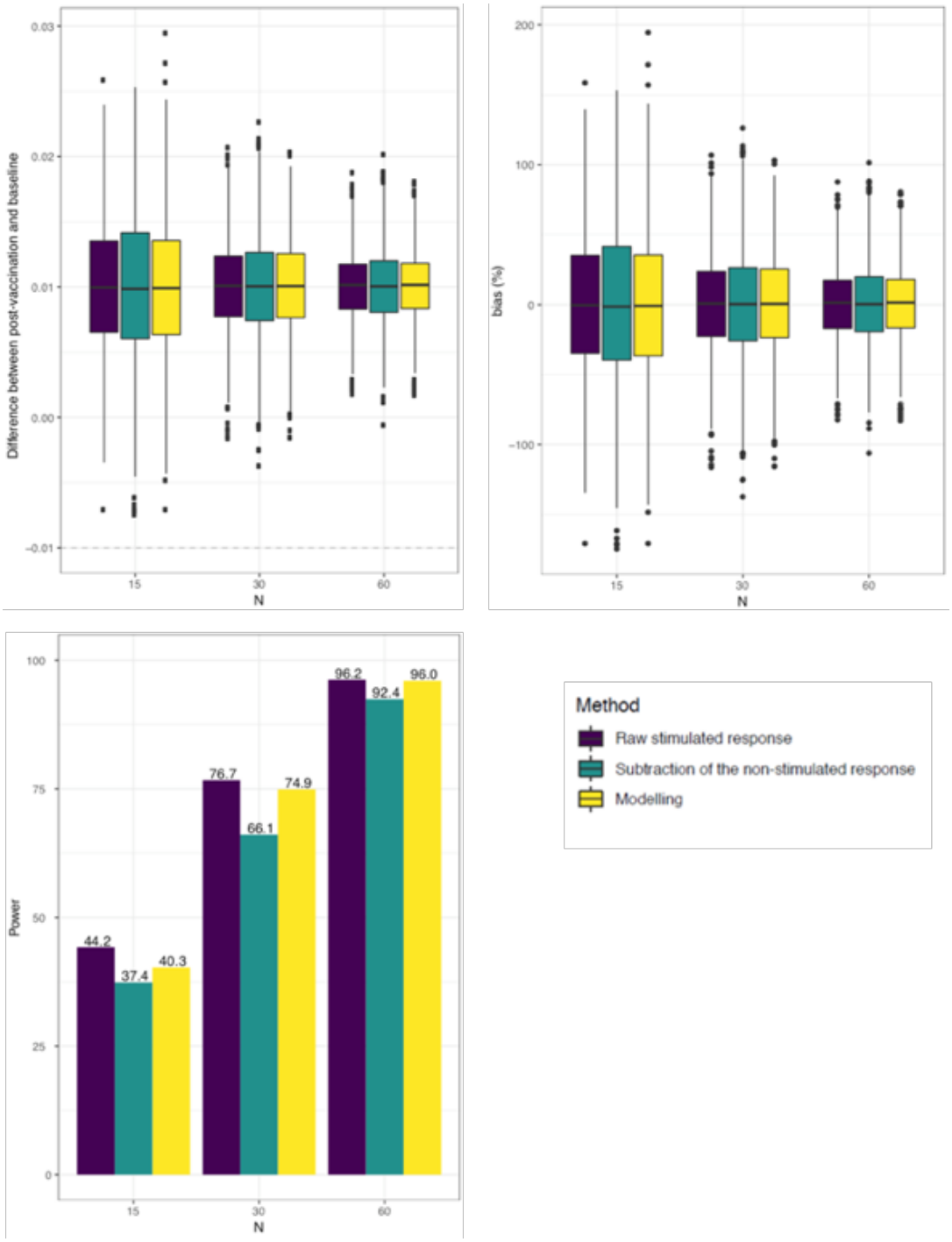
Appendix B – Figure 6. Scenario simulating a vaccine effect on stimulated response, Vaccine effect on non-stimulated response. Comparison of the difference between arm, the percentage of bias and the power between the modelling approach (model 2) and two conventional approaches (comparing

respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)

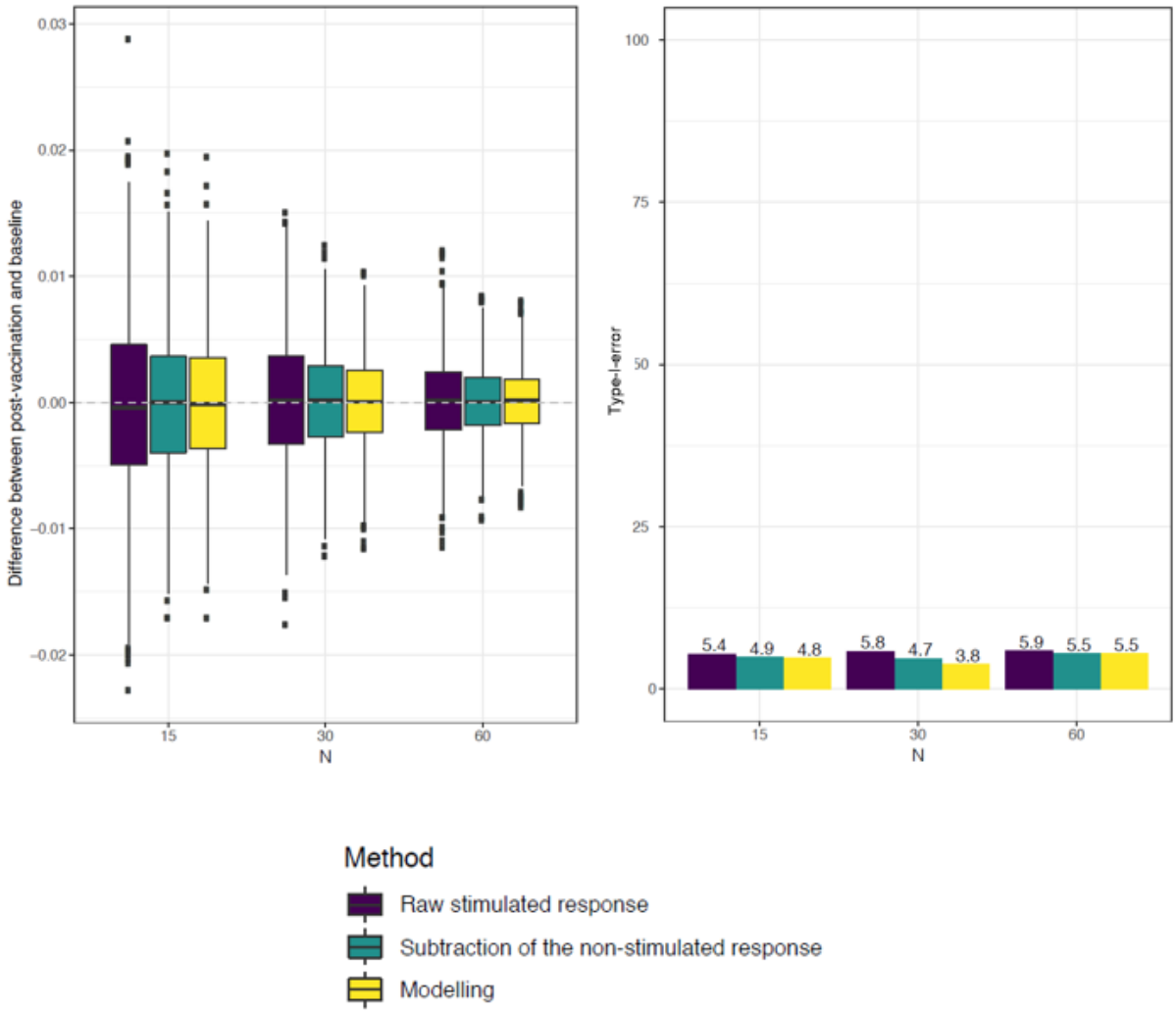
Simulation results – Model 2



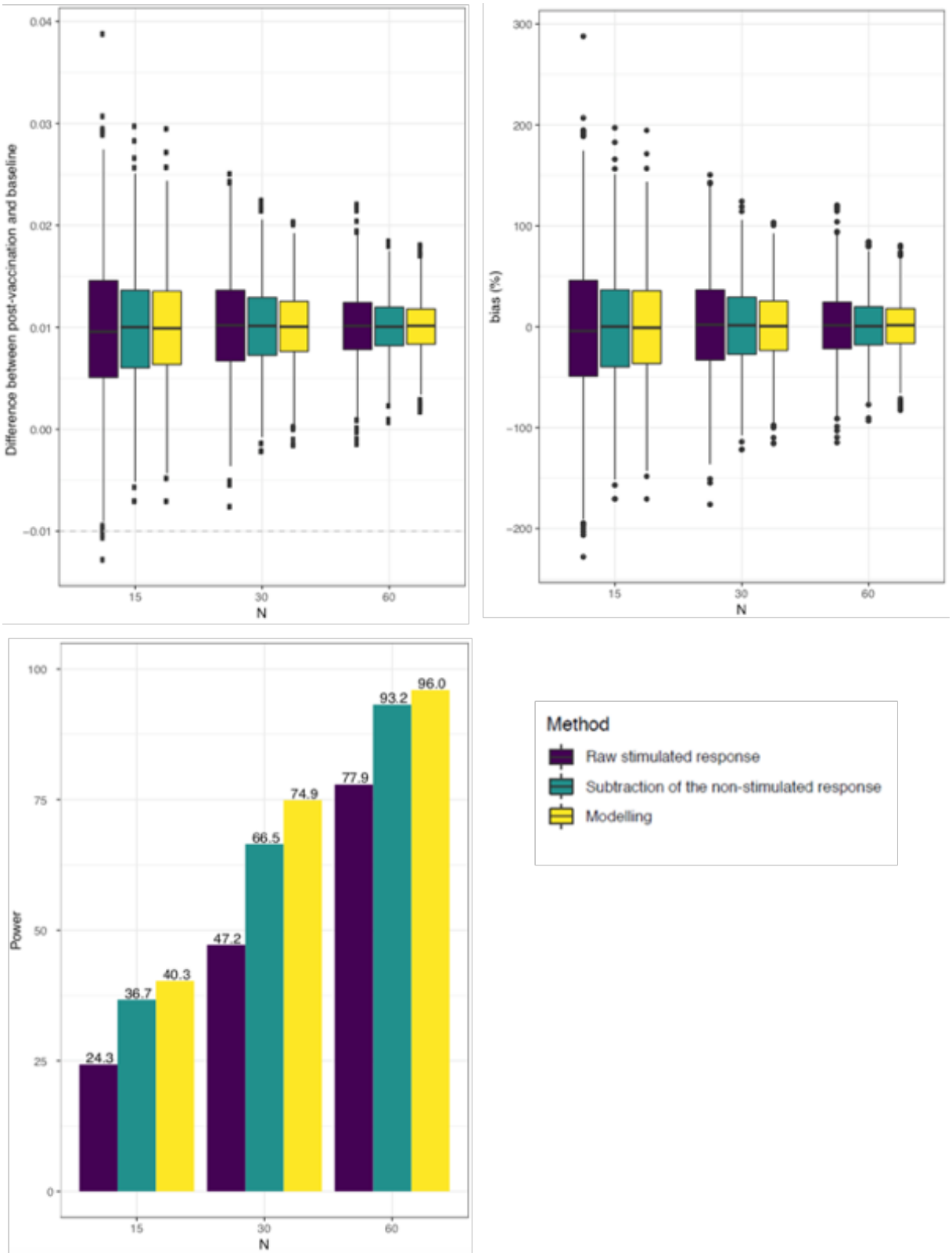
Appendix B – Figure 1. Scenario simulating a similar vaccine effect between arms on stimulated response. Comparison of the difference between arm (left) and the type I erorr risk (right) between the modelling approach (model 1) and two conventional approaches (comparing respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)



Appendix B – Figure 2. Scenario simulating a vaccine effect on stimulated response, no correlation between stimulated and non-stimulated response. Comparison of the difference between arm, the percentage of bias and the power between the modelling approach (model 1) and two conventional approaches (comparing respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)

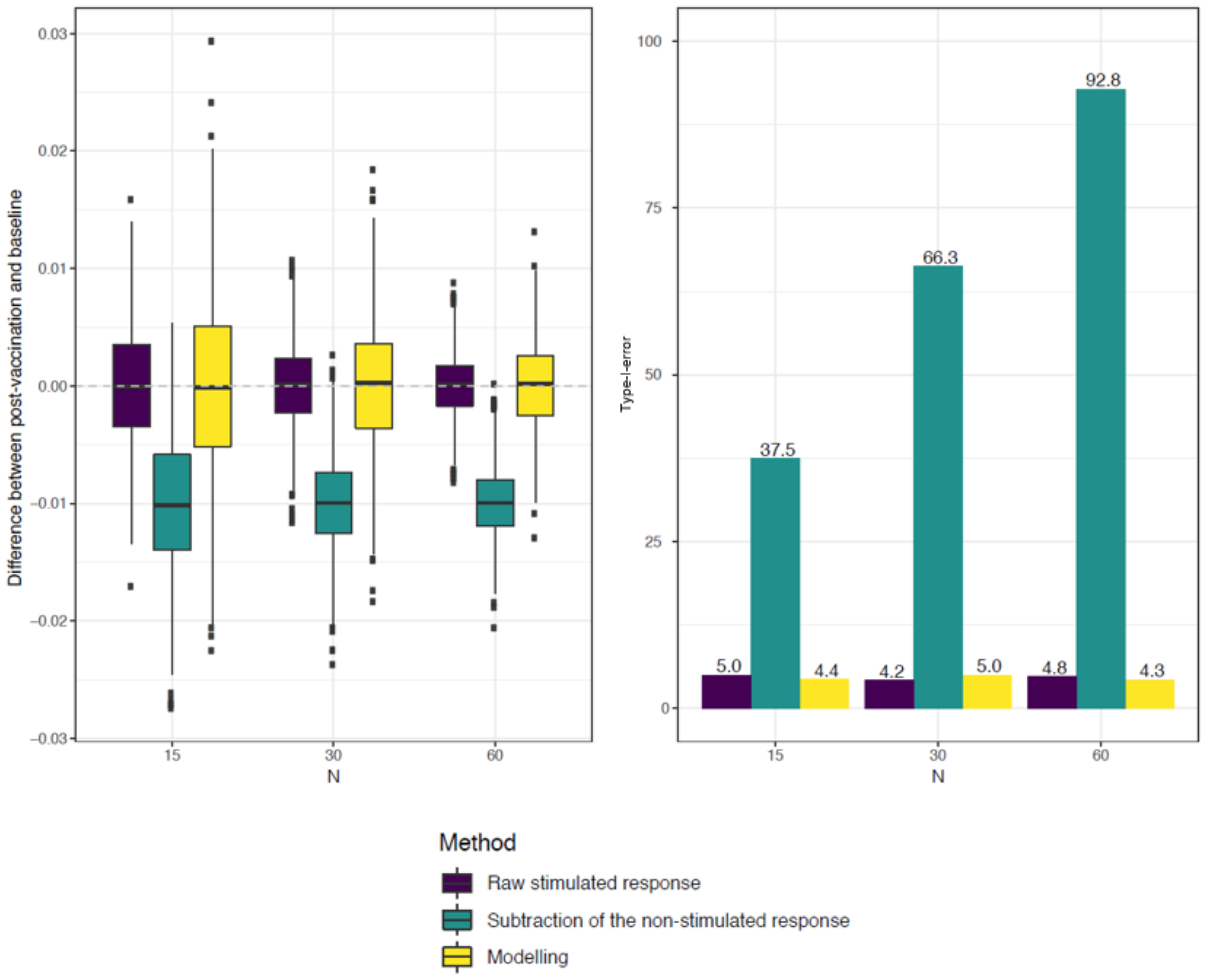


Appendix B – Figure 3. Scenario simulating a similar vaccine effect between arms on stimulated response, correlation between stimulated and non-stimulated response. Comparison of the difference between arm (left) and the type I error risk (right) between the modelling approach (model 1) and two conventional approaches (comparing respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)

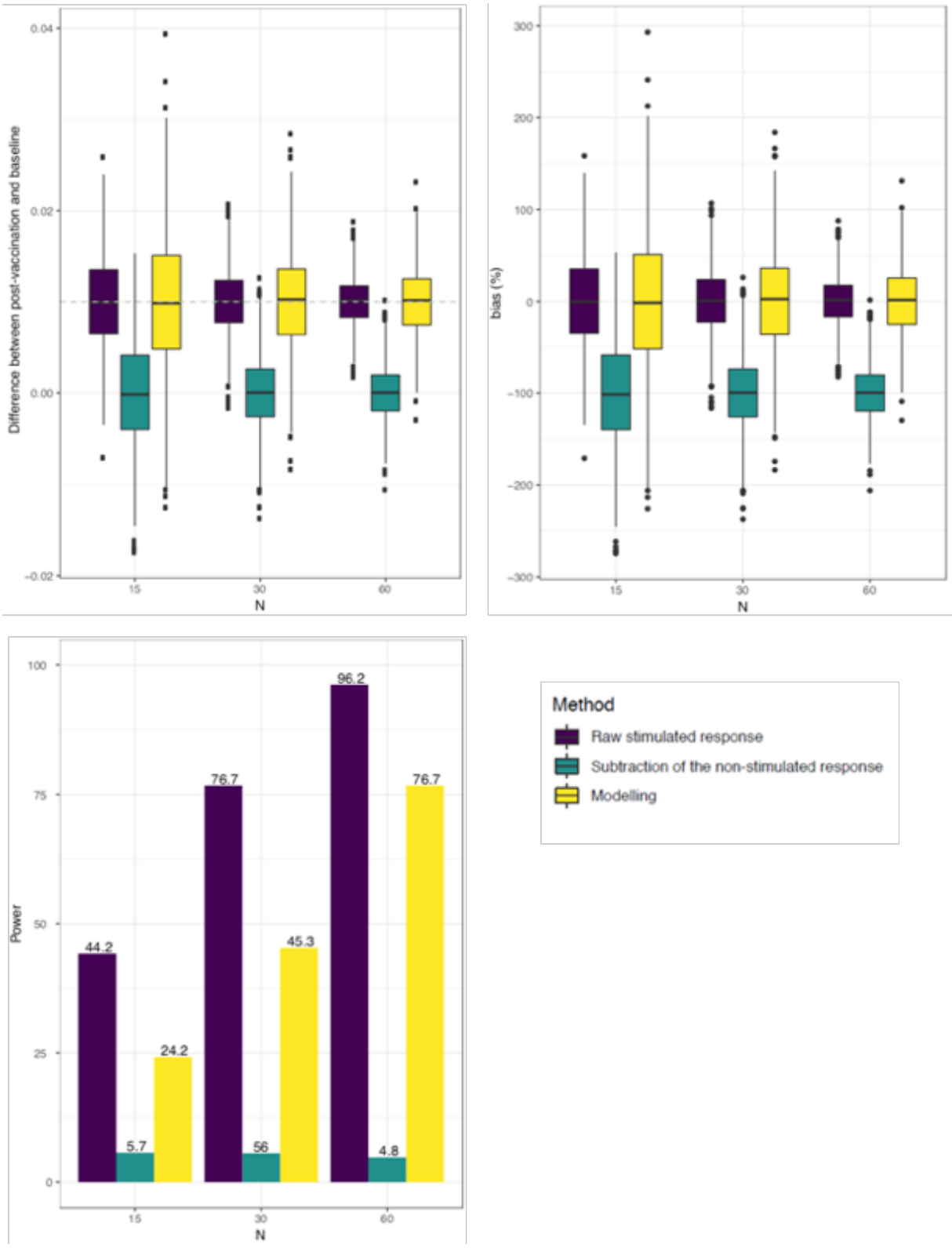


Appendix B – Figure 4. Scenario simulating a vaccine effect on stimulated response, no correlation between stimulated and non-stimulated response, correlation between stimulated and non-stimulated response. Comparison of the difference between arm, the percentage of bias and the power between the

modelling approach (model 1) and two conventional approaches (comparing respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)



Appendix B – Figure 5. Scenario simulating a similar vaccine effect between arms on stimulated response, Vaccine effect on non-stimulated response. Comparison of the difference between arm (left) and the type I error risk (right) between the modelling approach (model 1) and two conventional approaches (comparing respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)



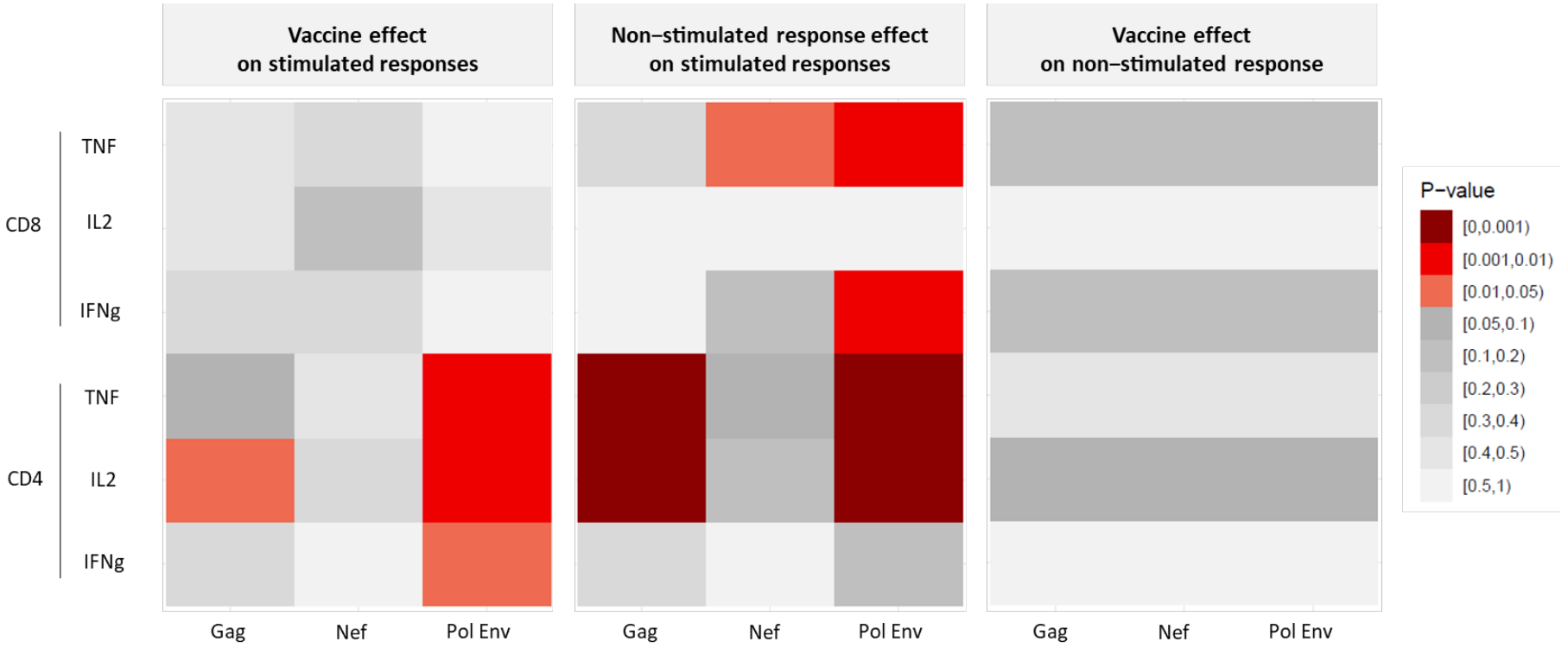
Appendix B – Figure 6. Scenario simulating a vaccine effect on stimulated response, Vaccine effect on non-stimulated response. Comparison of the difference between arm, the percentage of bias and the power between the modelling approach (model 1) and two conventional approaches (comparing

respectively the raw stimulated response or the stimulated response after subtraction of the non-stimulated response)

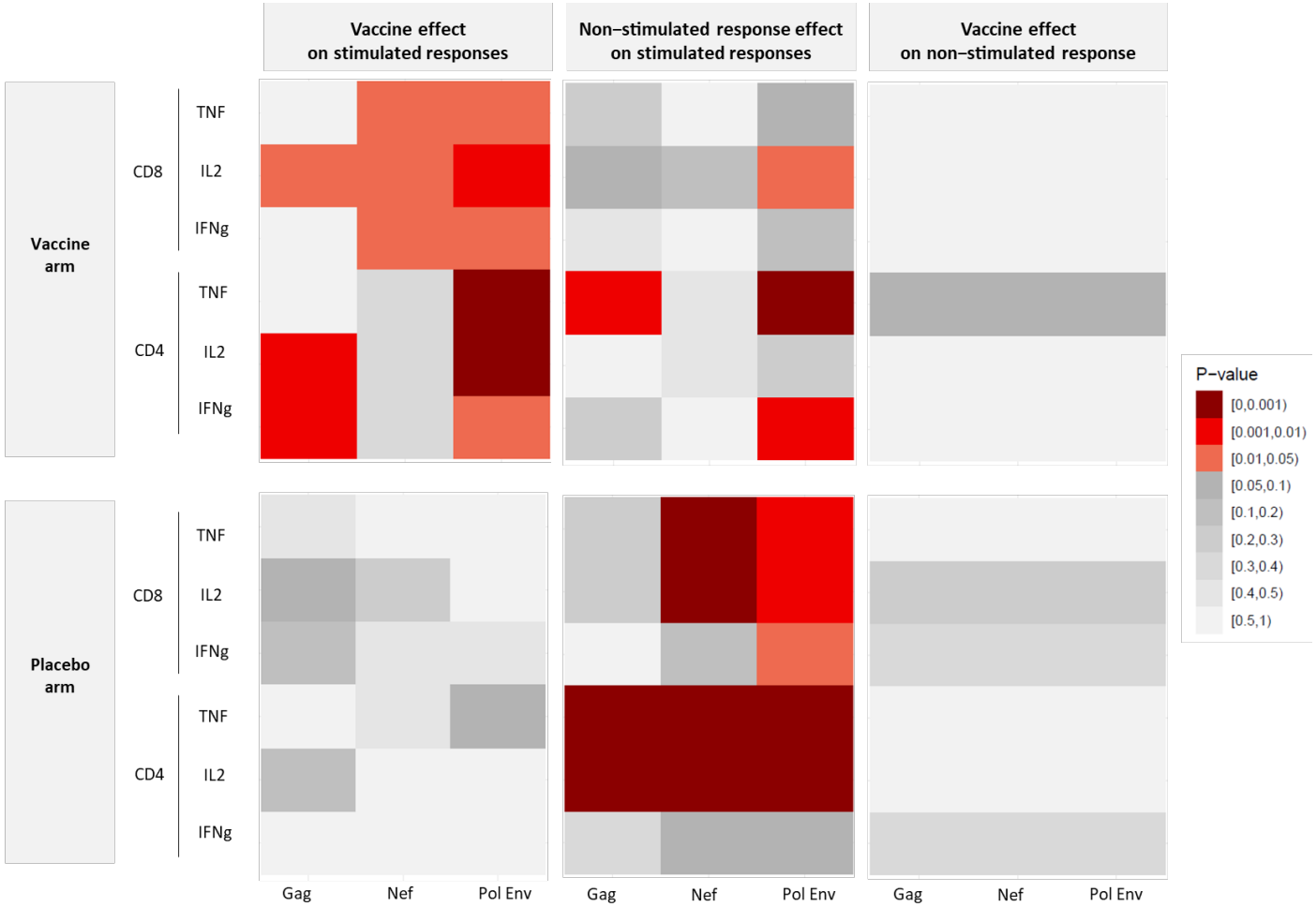
Appendix C – ICS analyzes

To assess antigen-specific T-cell responses, ICS assay was performed in a centralized laboratory (MIC-VRI, Creteil, France) on cryopreserved PBMC. PBMC were rested overnight and then stimulated (6h, 37°C, 5% CO2) with HIV peptide pools (1µg/ml) in the presence of co-stimulatoy molecules (anti-CD28 and anti-CD49d antibodies (1µg/ml)) and a protein transport inhibitor (Golgi Plug (1 µl/ml)) (BD Biosciences, Le Pont de Claix, France). SEB stimulation (100 ng/ml Staphylococcus Enterotoxin B; Sigma Aldrich, Saint Quentin Fallavier, France) served as positive control. After stimulation, cells were stained for dead cells with an amine-reactive dye (LIVE/DEAD Aqua, Invitrogen, Life Technologies, Saint Aubin, France) and with fluorochrome-conjugated monoclonal antibodies (anti-CD3 Alexa700, anti-CD4 PE, and anti-CD8 efluor 780; all from BD Biosciences) for 15 min at room temperature. After fixation and permeabilization using Cytofix/Cytoperm kit (BD Biosciences) for 20 min and staining with anti-IFN-γ PerCpCy5.5, -TNF-α PE-Cy7 and -IL-2 APC (all BD Biosciences) for 20 minutes at room temperature, PBMCs were re-suspended in Paraformaldéhyde 1% (BD Biosciences) and stored at 4°C until analysis. Data were acquired on a LSRII Fortessa 4-laser (488, 640, 561 and 405 nm) cytometer (BD Biosciences), analyzed using FlowJo software version 9.9.4 (Tree Star inc.). At least 250,000 events gated on $CD3^+$ were collected and analyzed using Boolean gating.

Appendix D – VRI02 Light trial

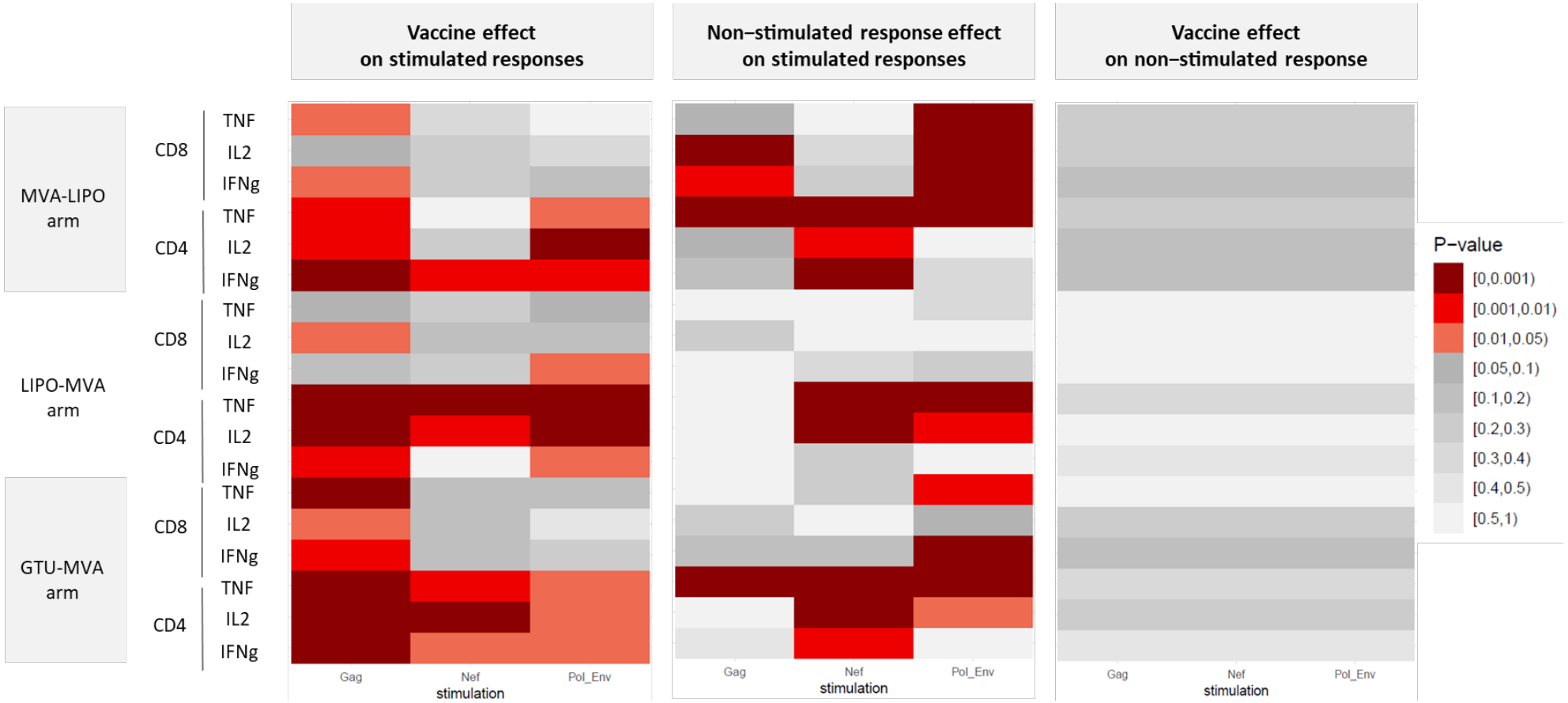


Appendix D – Figure 1. Description of the estimations obtained for inter-arm (vaccine versus placebo) comparison at W28 of the Light trial with modelling approach (Model 1)



Appendix D – Figure 2. Description of the estimations obtained for intra-arm (W0 versus W28) comparison of the Light trial with modelling approach (Model 2)

Appendix E – VRI01 trial



Appendix E – Figure 1. Description of the estimations obtained for intra-arm (W0 versus W30) comparison of the VRI01 trial with modelling approach (Model 2)